$2.95

Intelligence:

some recurring issues

Leona E. Tyler

VAN NOSTRAND INSIGHT SERIES

ABOUT THE EDITOR

LEONA E. TYLER (Ph.D., University of Minnesota) is Dean of the Graduate School at the University of Oregon. Prior to being appointed Dean she was for many years Professor of Psychology at the same university. Dean Tyler is the author of *The Psychology of Human Differences, The Work of the Counselor, Tests and Measurements, Developmental Psychology* (with F. L. Goodenough), and *Clinical Psychology* (with N. D. Sundberg). She also has contributed to various journals in the field, including *American Psychologist, Educational and Psychological Measurement,* and *Journal of Educational Psychology.*

2212
12.3.70

INTELLIGENCE

Some Recurring Issues

AN INSIGHT BOOK

INTELLIGENCE:

Some Recurring Issues

An Enduring Problem in Psychology

Edited by

LEONA E. TYLER

Dean, Graduate School
University of Oregon

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

VAN NOSTRAND REINHOLD COMPANY
NEW YORK CINCINNATI TORONTO LONDON MELBOURNE

S.C.E.R.T., West Bengal
Date 12-3-70
Acc. No. 2212

151
TYL

KRoy

Van Nostrand Reinhold Company Regional Offices:
Cincinnati, New York, Chicago, Millbrae, Dallas

Van Nostrand Reinhold Company Foreign Offices:
London, Toronto, Melbourne

Copyright © 1969 by LITTON EDUCATIONAL PUBLISHING, INC.

All rights reserved. No part of this work covered by the copyrights hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without written permission of the publisher. Manufactured in the United States of America.

Published by Van Nostrand Reinhold Company
450 West 33rd Street, New York, N.Y. 10001

Published simultaneously in Canada by
D. Van Nostrand Company (Canada), Ltd.

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Preface

The measurement of intelligence has been one of psychology's most outstanding accomplishments. Since 1905 when Binet and Simon first presented a graded series of intellectual tasks to be used in assessing a child's level of mental growth, hundreds of intelligence tests have been written and published for use in schools, employment offices, clinics, and the military services. From a practical point of view, they are a great success. In our complex, swiftly changing world, it is an advantage to be able to sort people out into groups differing in the kind of intellectual performance to be expected of them before they enter a course, a training program, or a job, instead of waiting for them to sort themselves out by the way they function in the situation. There is a constant and increasing demand. Intelligence tests sell.

From the beginning, however, intelligence testing has been the focus of criticism and controversy. Some of the criticisms result from misunderstandings about the purposes and limitations of psychological testing. Some of them come as a consequence of the misuse of tests by persons inadequately trained to select, administer, and interpret them. But some of them represent genuine and important psychological issues and value considerations that we should take seriously if we desire to improve our social institutions. The question of whether intelligence is a unitary trait or only a name for a combination of loosely related, separately developed aptitudes for special kinds of thinking, for example, is not a trivial one. What we think about this will influence our decisions about many other things, such as school policies, employment and unemployment, and the meaning of political equality.

Like so many psychological issues, these basic questions are never answered unambiguously and permanently. They keep cropping up in altered forms, decade after decade. But knowledge advances in ascending spirals, so that more penetrating and informative research investigations can be planned each time an old question recurs.

The purpose of this book is to bring together some definitive studies and statements on these recurring issues. The dates of the papers included range from the 1920's to the 1960's. With some notable exceptions, more emphasis has been given to the later rather than the earlier attempts to deal with the problems, because the later ones rest on a stronger foundation of research knowledge. Each of them, whether it came early or late in the history of the intelligence-testing movement, has been included because it brings to us something we should know or understand if we are to use intelligence tests intelligently in our efforts to build a rich and rewarding human society.

LEONA E. TYLER

Contents

Preface v

Part I—The Problem of Definition

1. INTELLIGENCE AND ITS MEASUREMENT: A SYMPOSIUM 3
E. L. Thorndike, L. M. Terman, F. N. Freeman, S. S. Colvin, Rudolf Pintner, S. L. Pressey, V. A. C. Henmon, Joseph Peterson, L. L. Thurstone, Herbert Woodrow, M. E. Haggerty

2. INTELLIGENCE AS THE TESTS TEST IT 25
Edwin G. Boring

Part II—Intelligence and Learning Ability

3. INTELLIGENCE AND IMPROVEMENT IN SCHOOL SUBJECTS 31
Herbert Woodrow

4. INTELLIGENCE TEST SCORES AS INDICATIVE OF ABILITY TO LEARN 41
J. W. Tilton

5. STUDIES IN CONCEPT ATTAINMENT: EFFECT OF INSTRUCTION AT TWO LEVELS OF INTELLIGENCE 48
S. F. Osler and S. R. Weiss

Part III—INTELLIGENCE AND TEMPERAMENT

6. COGNITIVE, CONATIVE, AND NON-INTELLECTIVE INTELLIGENCE 61
David Wechsler

7. A NONINTELLECTUAL INTELLIGENCE TEST 75
Harrison Gough

Part IV—INTELLIGENCE OR INTELLIGENCES?

8. THE ABILITIES OF MAN 89
Charles Spearman

9. THREE FACES OF INTELLECT 98
J. P. Guilford

10. THE STUDY OF GIFTEDNESS: A MULTIDIMENSIONAL APPROACH 121
J. W. Getzels and P. W. Jackson

Part V—THE GROWTH AND DECLINE OF INTELLIGENCE

11. ON THE GROWTH OF INTELLIGENCE 141
Nancy Bayley

Part VI—THE NATURE-NURTURE CONTROVERSY

12. SCHOOLING MAKES A DIFFERENCE 171
Irving Lorge

13. DEPRIVATION, DEVELOPMENT, AND DIFFUSION 187
Susan W. Gray and Rupert A. Klaus

14. THE INHERITANCE OF MENTAL ABILITY 207
Cyril Burt

15. HEREDITY, ENVIRONMENT, AND THE QUESTION "HOW" 237
Anne Anastasi

Part VII—INTELLIGENCE TESTING AND ITS CRITICS

16. LOST: OUR INTELLIGENCE? WHY? 259
Quinn McNemar

Part I

THE PROBLEM OF DEFINITION

1

Intelligence and Its Measurement: A Symposium

When psychologists first began to grapple with the new task of measuring intelligence, they assumed that everyone agreed about what was meant by the term. But it soon became apparent that there was no general agreement at all about this. In 1921 a very influential symposium brought together the views of fourteen of the leading investigators of the time. The portions of this symposium included in the following pages highlight the differences in definition and call attention to the major problems with which the testing movement was to struggle in the decades ahead.

The first selection, then, is an abridged version of "Intelligence and Its Measurement: A Symposium" from the Journal of Educational Psychology, *1921,* 12, *123–147 and 195–216.*

Editorial introduction: Probably the most striking advance of our generation in the practical application of psychological technic to educational and social affairs is the movement for the definition and measurement of intelligence. Especially impressive is the widespread effort now devoted to the construction and use of individual and group tests. It appears that this is an opportunity for this journal to serve as an effective clearing house for mature opinion on a most important problem. Accordingly, we have asked 17 leading investigators to contribute to a symposium on the following topics:

"(1) What I conceive 'intelligence' to be, and by what means it can best be measured by group tests. (For example, should the material call into play analytical and higher thought processes? Or, should it deal equally or more considerably with simple, associative, and perceptual processes, etc.?)"

"(2) What are the most crucial 'next steps' in research?"

Those invited to take part were: Doctors Bell, Buckingham, Colvin, Dearborn, Freeman, Haggerty, Henmon, Peterson, Pintner, Pressey, Ruml,* Terman, Thorndike, Thurstone, Whipple, Woodrow, Yerkes.

Of these, all but three have signified their desire to contribute material. We print in this issue statements by Drs. Thorndike, Terman, Freeman, Colvin, Pintner, Ruml, and Pressey. Contributions from the other workers will be published in the April issue.

The editors hope that a vigorous discussion will grow out of these initial statements and will present further phases of the matter in subsequent issues of the *Journal.*

I. By E. L. THORNDIKE

Teachers College, Columbia University . . .

1. The nature and measurement of intelligence. If we inventory the behavior of men and try to divide it up according as intellect, character, skill, taste or temperament is primarily involved, we shall agree fairly well in, say, ninety per cent of the cases. If, however, we try to make the division absolute we may agree very seldom. It is probably unwise to spend much time in attempts to separate off sharply certain qualities of man, as his intelligence, from such emotional and vocational qualities as his interest in mental activity, carefulness, determination to respond effectively, persistence in his efforts to do so; or from his amount of knowledge; or from his moral or esthetic tastes. Even so apparently remote a trait as muscular strength may in some cases cooperate almost indistinguishably in the production of what we would all call intellectual products. Thus a great scholar's achievement may be in part due to eye muscles which help make reading a pastime.

Taking these cases of behavior which are generally accepted as matters of intellect and trying to place each as primarily a matter of response to situations directly sensed, or as primarily a matter of planning, we shall again agree fairly well. So also

* The statements of Ruml (VI) and Dearborn (XII) have been omitted.—ED.

if we rate them as primarily responses to concrete particulars or primarily responses to abstract qualities and relations. It would however be difficult and probably unwise to try to separate off *sharply* the responses concerned with directly sensed situations from responses concerned with planning; or those concerned with concrete features of things and men from those concerned with ideas and symbols. Hoeing corn and shooting a rabbit are easily distinguished from studying botany and ballistics, but behavior shows all sorts of intermediate forms.

Realizing that definitions and distinctions are pragmatic, we may then define intellect in general as *the power of good responses from the point of view of truth or fact*, and may separate it according as the situation is taken in gross or abstractly and also according as it is experienced directly or thought of. The power of good responses to abstract qualities and relations rather than gross total facts and to ideas rather than direct experiences may be called the more intellectual variety of intellect.

Either variety is, as human beings now are, specialized further according to the data operated on in the response and according to the form of the operation. The goodness of the response in any individual varies according to the particular task. The child who is the best of a thousand of his age at the undoubtedly intellectual task of mental multiplication of two-place numbers will not be the best at the equally indubitably intellectual task of thinking out verbal puzzles. As we change the task from space relations to animals, to plants, to machines, to moral issues, to numbers, the correlation never holds up to 1.00. As we change from accuracy in perception, to accuracy in memory, to accuracy in inference, all with the same data, the correlation again fails to hold at 1.00.

A part of this break to below perfect correlation is presumably due to circumstances of life and training which have given unlike amounts of emphasis to different data and to different forms of operation in the case of each of the thousand. But if a thousand were taken who had had identical training, it seems certain that the specialization of intelligence would still be found, the correlations still failing to be unanimously 1.00.

In measuring a person's general status in intelligence and in inferring therefrom what his rank in native intellectual capacity in general is, what we do is to test him with a fair sampling of data and operations. If his opportunities of training in respect to these have been inferior or superior to the group with whom he is to be compared we make the necessary allowance. This sampling should be wide enough and its various components should be easily enough weighted, so that the resulting judgment should be about his *general* status and *general* capacity—if we are to claim that it is general.

Some of us have, I fear, claimed a generality for our measures of status and a surety of inference from them to original inborn capacity which it would be very hard to justify. The estimates which the psychologist makes with his tests are much better than those which parents or teachers or ordinary medical practitioners make of the same facts, so that we are justly proud of them, but we should be the first to recognize their limitations. The value of a test score is its value in prophesying how well a person will do in other intellectual tasks. Our claims may wisely be limited to the actual demonstrated power of prophecy. For example, consider a score attained by a 12-year-old boy in say a combination of Stanford Binet, National A and B and Haggerty Delta 2 (two trials of each). If the boy has had ordinary American opportunities, this score will prophesy rather accurately how well he will respond to intellectual demands in the cases of "book-learning" at the time and for some years thereafter, and very possibly for all his life. It will prophesy less accurately how well he will respond in thinking about a machine that he tends, crops that he grows, merchandise that he buys and sells and other concrete realities that he encounters in the laboratory, field, shop and office. It may prophesy still less accurately how well he will succeed in thinking about people and their passions and in responding to these. We know that, taking people as we find them, the ability measured by verbal tests is not the same as the ability measured by non-verbal tests; and there is reason to expect other similar specializations.

The intercorrelations of these various "intelligences" are, of course, high enough to make a measure of any one of them a better index of any other than the average parental hope or teacher's opinion is likely to be. If our 12-year-old boy is at the five-percentile station in our tests it will rarely happen that he will rank above average intelligence over any large area of mental activity; save by the drive of a great interest and the expense of much time upon the activity in question. On the other hand to assume that we have measured some general power which resides in him and determines his ability in every variety of intellectual task in its entirety is to fly directly in the face of all that is known about the organization of intellect.

The relative weight to be attached to analytical and selective and to perceptual and simple associative processes can only be decided by the correlations and partial correlations of these with the criterion which the test score is to prophesy, or some experimental facts giving equivalent information, considered with due reference to the time-cost of making the prophecy. In general, tasks which require efficiency in analyzing a situation into elements, selecting and weighting elements to fit a problem and organization or thinking many things together seem to give much better results per dollar or hour of cost. They perhaps include and sum up the action of many simpler associative processes. But a straightforward information test is also a valuable element. If time-cost is disregarded, a sampling of every kind of intellectual operation, and with every kind of *datum,* will, if properly weighted, improve the prophecy of *general* intelligence. Unless it is properly weighted, however, it may injure it.

2. *Next steps in research.* Having used up my allotment of space in the previous discussion, I can only note that more research into every feature of intelligence and its measurement is needed, especially researches on the intercorrelations "total" and "partial" of the various forms of intellectual work of the world; on the intercorrelations "total" and "partial" of the abilities used and usable in our tests; on the permanence of our IQ's and EQ's; and on the "total" and "partial" correlations of each test at each age with itself at later ages and with each feature of the intellectual work of the world. The form of distribution of intelligence in the general adult population is also a matter of great practical moment. In relation to educational and industrial uses the problem of the effect of "coaching," special and general, is fundamental.

II. By L. M. TERMAN

Leland Stanford University

1. The nature and measurement of intelligence. Meumann has pointed out that the fault of Stern's teleological definition of intelligence as "general adaptability to the new problems and conditions of life," lies in the fact that it furnishes no clue for judging the value of different kinds of adaptation. Meumann would reverse Stern's procedure by first finding out what is demanded of intelligence and then analyzing the mental functions which meet that demand. In my opinion this is the only method of approach which will bring us any nearer to a psychological solution of the intelligence problem.

If we accept this view it is evident that the important intellectual differences among men will not be found on the sensory, perceptual, or purely reproductive level. It is well known that a moron may be able to see, hear, taste or smell, react to a signal, balance a bicycle, steer an automobile, or cancel A's about as well as an intellectual genius. The latter would be somewhat his superior in memory for non-sense syllables, would excel him still more in logical memory, and would outclass him hopelessly in the ability to distil meanings from the raw products of sensation and memory. The essential difference, therefore, is in the capacity to form concepts to relate in diverse ways, and to grasp their significance. *An individual is intelligent in proportion as he is able to carry on abstract thinking.*

One may, of course, question our grounds for designating any kind of mental activity as "higher" or "lower" than another. Why, it may be asked, should certain types of mental processes be singled out for special worship? In fact, it is frequently intimated that the individual who flounders in abstractions but is able to handle tools skillfully, or play a good game of baseball, is not to be considered necessarily as *less* intelligent than the individual who can solve mathematical equations, acquire a huge vocabulary, or write poetry. The implication is that the two individuals differ merely in having different *kinds* of intelligence, neither of which is higher or better than the other.

It is difficult to argue with anyone whose sense of psychological values is disturbed to this extent. Such an individualistic view is not disposed of by calling attention to the obvious fact that civilization, with its science, art, government, religion, philosophy, and systems of credit, is unthinkable except as a product of concept elaboration and symbolic thinking; our opponent can retort that it is only our intellectual snobbery which leads us to regard the state of so-called civilization as "higher" than that of primitive man!

It can not be disputed, however, that in the long run it is the races which excel in abstract thinking that eat while others starve, survive epidemics, master new continents, conquer time and space, and substitute religion for magic, science for taboos and justice for revenge. The races which excel in conceptual thinking could, if they wished, quickly exterminate or enslave all the races notably their inferiors in this respect. Any given society is ruled, led, or at least molded by the five or ten per cent of its members whose behavior is governed by ideas. The typical pick-and-shovel man does his thinking chiefly on the sensori-motor and perceptual levels. Add a little more ability to think on the representative level and he may be able to repair your automobile, build you a house according to an architect's specifications, or nurse you in illness. Add a large measure of ability to associate abstract ideas into complex systems and he can design a new type of engine, draft the plans for a skyscraper, or discover a curative serum.

What I have said may seem like an over-elaboration of the obvious. Such a discussion ought, indeed, to be unecessary. Unfortunately it is not. Many criticisms of the current methods of testing intelligence rest plainly on a psychology which fails to distinguish the levels of intellectual functioning or to assign to conceptual thinking the place that belongs to it in the hierarchy of intelligence. If an intelligence test can be shown to depend upon the language factor (i.e., upon the ability to think in terms of symbols), this is sufficient in the eyes of some psychologists to condemn it as non-valid. The subject who can not acquire a normal vocabulary, see the point of a fable, or be taught to read a paragraph with understanding is considered to have demonstrated his intelligence if he can trace a simple maze or assemble the fragments of a formboard. In fact the "idea thinker" is sometimes spoken of disparagingly or even a little contemptuously, particularly in the case of a child whose superior ability in this respect places him in conspicuous contrast with other children of the same age.

But if intelligence is the ability to think in terms of abstract ideas, we should expect the most successful intelligence tests to be just those which involve the use of language or other symbols. We should also be justified in demanding that an intelligence test should correlate well with what we may call "school educability." As a matter of fact it is precisely tests of this type which are surviving in the struggle for existence: tests involving arithmetical reasoning, language completion, naming opposites, matching proverbs, completing analogies, understanding difficult passages, etc. The list could of course be greatly extended, but all tests which are notably successful in measuring intellectual differences among adults have something in common with those named, even when they work with concrete materials instead of with words. . . .

III. By F. N. FREEMAN

University of Chicago

1. The nature and measurement of intelligence. I conceive intelligence to be a somewhat more inclusive capacity than is implied when it is used as a name for our present tests. For this reason, it seems to me that it would be better to use a term of somewhat narrower connotation to designate these tests. The mental capacity designated by the term intelligence seems to me to include, besides the elements which are usually measured by our tests, certain other types of capacity which they measure scarcely at all.

The capacities measured by our tests may be described from the structural point of view as involving chiefly: sensory capacity; capacity for perceptual recognition; quickness, range of flexibility of association; facility in imagination; span or steadiness of attention; quickness or alertness in response. Excellence in capacities of this sort is well designated by the term brightness. We recognize, however, that a person's achievement, or even his intellectual capacity, does not seem to correspond perfectly with his capacity in any of these various specific traits or with all of them put together. We sometimes say that a person is very bright but not very intelligent. We mean that he does not use his mental powers in such a way as to be most productive. The characteristic which I am referring to is sometimes called temperament or moral character, but it seems to me it can be shown that there are certain intellectual traits which are left out of our present scheme of tests and which should be included in a total conception of intelligence.

These additional characteristics may be described as: mental balance; co-ordination of the mental processes; the judicious management of the processes of learning or reflection; mental control; mental adjustment; the direction of the attention toward the significant aspects of experience; a due degree of non-suggestibility; the adoption of intellectual purposes and the adaptation of means to their satisfaction; sensitiveness to significant combinations between experiences which illuminate one another or which are effective in building up systems of thought; balanced and sane reaction to the entire world of things, ideas and persons. . . .

IV. By S. S. COLVIN

Brown University

1. Nature and measurement of general intelligence. General intelligence has been defined as "general mental adaptability to new problems and conditions of life." To my mind this definition is somewhat too narrow. In a very true sense intelligence is mental adaptability to environment. This conception, however, is in one respect too broad since it includes instinctive adaptations as well as those that have been acquired through experience. Of course, psychologists sometimes speak of the psychic life of micro-organisms and frequently use the term intelligence in connection with instinctive acts of such animals as ants, bees and wasps, whose adaptations to environment seem to be almost entirely on the plane of instinct. On the whole, I consider the most helpful viewpoint from which to consider intelligence is that it is equivalent to the capacity to learn. *An individual possesses intelligence in so far as he has learned, or can learn to adjust himself to his environment.* In a sense this conception is substantially the same as that quoted in the definition first given. However, it does not unduly emphasize the problem aspect of intelligence and rightfully attributes intelligence to those animals whose sole ability to learn is confined to the hit-and-miss try-out of experience ("trial and error").

Psychologists have accepted this definition practically if not theoretically. An inspection of intelligence examinations clearly shows that those who framed them have not confined their tests to problem solving, even in its rudimentary forms. These tests measure an individual's intelligence largely in terms of what he has learned, thus obtaining indirectly a measure of his learning ability. Vocabulary tests, range of information tests, "same and opposites" tests, tests of fundamental operations in arithmetic, and the like, call for little ingenuity. If the individual has the requisite skill and knowledge he can satisfactorily perform these tests. They are appropriate tests for intelligence only on the theory that they test ability to learn by discovering what has already been learned. Even those tests that demand verbal and mechanical ingenuity are valid only in so far as individuals taking these tests have had common opportunities to learn the elements necessary in solving the problems involved in sentence completion, thought interpretation and the like. It must be remembered that even the ability to think in a sustained and logical manner is based on having learned how to think. Thought is a habit and is acquired through learning. In a word, the validity of all mental testing rests on the fundamental assumption that those tested have had a common opportunity to learn the skills, facts, principles and methods of procedure exemplified in the tests. It follows for this reason that many of the standard school tests now in use are reasonably good measures not only of specific aspects of acquired intelligence, but also of general (innate) intelligence. All the individuals in the school group tested have had the same training, or at least very similar training. Some have learned more, others, less; those who have learned less possess less learning capacity,—hence less general intelligence. . . .

V. By RUDOLF PINTNER

The Ohio State University

1. Nature and measurement of intelligence. I have always thought of intelligence as the ability of the individual to adapt himself adequately to relatively new situations in life. It seems to include the capacity for getting along well in all sorts of situations. This implies ease and rapidity in making adjustments and, hence, ease in breaking old habits and in forming new ones. Fundamentally, this leads us back to the general modifiability of the nervous system. An organism whose nervous system is very modifiable can adjust itself quickly to new types of situations and in this way react with more intelligence to a greater number of situations.

The reactions to which we apply the terms "intelligent" or "unintelligent" are differentiated from those to which we apply the term "emotional," although any sharp division between these two types can hardly be made. Nevertheless, the distinction is a practical one which it is useful to retain. . . .

Because we are dealing with something which we hardly know how to define, because we are groping for means to measure reactions to extremely complex and different situations, we must be very careful not to limit in any way the type of material we use. Intelligence is shown in dealing with things as well as with words, in dealing with living men and women as well as with symbols, in handling a paint brush or chisel as well as a pencil, in carving stone or laying brick as well as in making verbal responses. We, therefore, need intelligence tests of all types. And by intelligence tests I mean tests of the general ability to do all sorts of things as opposed to educational and trade tests which are specifically made to measure the knowledge which an individual has been directly taught. If we believe in this large view of intelligence, we are painfully aware of the paucity of tests that exist at the present time. I cannot believe that we have exhausted our ingenuity in the construction of different types of tests and I hope to see many new types of tests constructed in the future. All sorts of tests and various combinations of all sorts of tests are needed. . . .

VII. By S. L. PRESSEY

University of Indiana

1. Intelligence and its measurement. The writer has been asked for a statement as to what he conceives intelligence to be, and as to what types of test materials can best be used for the measurement of intelligence. Frankly, he is not very much interested in the question—although a large part of his time goes to work with tests of "intelligence." Instead, he is interested to know what such tests will *do,* in solving this or that problem. Suppose, for instance, one is intending to survey a school system in order to find those children who should be put into special classes—classes for those so dull as to be incapable of profiting by the regular school work. Tests in the school subjects cannot be used for this purpose, since some children would then test low because they had had their first training in poor country schools, or had been out of school because of poor health. Under these circumstances, one naturally makes up an examination not involving matters taught specifically in school, but a test as much like school work in other respects as possible. We have, then, a "general" test.

After a number of such tests have been made up the next step is, naturally, to evaluate them according to the efficiency with which they deal with the practical problem for the solution of which they were made. The obvious procedure is to give the tests to (1) a group of children who are known, from long and intensive study of them, to belong in special classes, and (2) another group known to belong in the regular classes. The tests which separate these two groups most distinctly are, then, the tests to include in the final scale.

Measurement of capacity for school work by such a paraphrase of school work, but with the specific elements of schooling left out, is surely a very natural thing to do. It is also natural to attempt the selection of children at the other extreme—the exceptionally bright children, who are capable of more advanced work than they are now doing—by means of such tests. The capacity of these children for this advanced work cannot be determined by giving them some of this advanced work to do; the point of the whole situation is that opportunity has not been given to acquire the specific elements in this advanced work. In fact, *any* effort to get back of the school's actual disposition of its children, into classes and grades, naturally suggests the use of such paraphrase of school work, but with the specific elements eliminated. And the usefulness of different types of this material must be determined by actual trial with the concrete problem to be dealt with.

Well—and if such tests are of service in dealing with a variety of problems one may come to think of them as tests of a fundamental, unitary, general intelligence. An elaborate theory, with elaborate statistical expression, may then be developed. But the hypothetical nature of the concept must be kept in mind (and the educational situation which first gave rise to such tests, also). Particularly, one must not permit himself to be hypnotized by the statistics. Always, it must be kept in mind that there is no value in measuring intelligence simply for the sake of measuring it; the tests are given in an effort to deal with some practical problem. And it is in proportion as they deal efficiently with that problem that they are to be considered satisfactory.

Frankly, then, the concept of general intelligence is for the writer, primarily, simply a working hypothesis which has been very helpful in the first attack upon problems of prognosis. It should continue as a working hypothesis until other hypotheses are found to work better. But investigators must not be so dominated by the concept of general intelligence that they fail seriously to consider other possibilities. Particularly to be avoided is the easy notion that a test of intelligence will solve all educational difficulties, that intelligence testing is the longed for short-cut method which will put everything in the schools finally and completely to rights. The theory is simply an assumption that prognosis problems can be generalized. It has been of great use in suggesting certain general lines of approach—and in giving us courage to attack the otherwise impossibly complex prognosis problem. It will do equally great harm if it prevents analysis of the prognosis problem, now that methods are becoming sufficiently refined to make this possible. The safest, most empirical, and most practical method of making this analysis (to the writer's mind) is to study the relation of the tests to particular practical problems. From such study any general factors will, in time, emerge of themselves.

VIII. By V. A. C. HENMON

University of Wisconsin

1. The nature and measurement of intelligence. Intelligence, in the ordinary acceptance of the term, has been defined by Lester F. Ward to be "intellect coupled with the product of its operation," or in other words, "intelligence is intellect *plus* knowledge." This certainly corresponds to the use of the term in pedagogical and sociological theory and in practical life. The intelligent man is the well informed man and one who is capable of readily appropriating information or knowledge. Without doing violence to language the psychologist can not give it any other technical significance. Intelligence, then, involves two factors,—the capacity for knowledge and knowledge possessed. The untutored savage or barbarian may have high intellectual capacity, but without knowledge we should not ordinarily call him an intelligent man. We could best only say that he had a fine mind, a fine intellect, high intellectual power, or else that he possessed a high native intelligence. But native intelligence from the point of view of this definition is a misnomer, as is also the customary distinction between intelligence tests and achievement tests. Nothing but confusion results from using the term intelligence to designate the level reached in the acquisition of knowledge or mental age, in which sense it is something which can be developed and trained, and using the same term to designate intellectual capacity or mental alertness which is probably given once for all with one's constitution and can not be materially changed by anything we do. Achievement tests while measures of knowledge possessed are in turn indices of capacity for knowledge, though valueless for comparative purposes of crude, uninstructed intellect, only because there has been no guarantee of equality of opportunity. Intelligence tests, on the other hand, as at present constituted, are a mixture in varying proportions of tests of knowledge, for example, the constantly recurring tests of arithmetical reasoning and of information, with those whose novelty is such that specific past experiences function much less readily and there is an approximation toward relative equality of opportunity. In any case, both groups of tests are measures of both native and acquired intelligence with differences in emphasis. The high correlations between intelligence tests and achievement tests and the high correlations between intelligence tests and teachers' estimates or scholastic standings indicate their essential similarity. Teachers' estimates and school marks, which are used as measures of reliability and validity of intelligence scales, in default of any absolute standard of reference, are based on what children know and on their capacity to know.

Intelligence ought to retain this broad significance. To define it with Stern as "a general capacity of an individual consciously to adjust his thinking to new requirements," which is substantially James' definition of reasoning, or with Meumann as a general capacity for "independence, originality and productiveness in thinking" is to narrow its meaning arbitrarily and to neglect its compound nature. The scholarly or erudite man who has merely acquired the knowledge created by others may not represent as high a degree of intelligence as one who is independent, original and productive in his thinking, but we should scarcely say that he is unintelligent. Intelligence is indicated by the capacity to appropriate truth and fact as well as by the capacity to discover them.

What we want, of course, is a measure of crude intellect or mental alertness as such, a scale for intellect not a scale for intelligence. Any test which requires mental processes, be they either on the sensory-motor level or on higher levels, which measures the sensitiveness, responsiveness, and retentiveness of the nervous system, is a test of intellectual capacity. It is only because of the absence of an absolute standard of reference, the inability to discount the effect of past experiences, and the narrow conception of the meaning of intelligence, that the tests of higher mental processes appear to be better measures of intellect, and not because they involve intellectual capacity any more evidently. The essential thing in a test of pure intellect is to secure such novelty as to reduce inequality in opportunity and training to a minimum, and such complexity as to reveal differences clearly and unambiguously. It is not a matter of the kind of mental process involved primarily. A properly constituted test of intellect ought, therefore, to take a wide sampling of so-called lower as well as higher processes and the relative weights to be assigned to each element determined. The tests must be extended to measure the capacity to learn in other directions than the ability to use words, signs, and symbols. The so-called general intelligence tests are not general intelligence tests at all but tests of the special intelligence upon which the school puts a premium. The extension of the term intelligence to include what Thorndike has called mechanical intelligence, the ability to manipulate things, and social intelligence, the ability to manage men, means that we need not one, but several scales of intellect from which we can secure an intellect profile. We have been gradually evolving a type of test for abstract intellect that is of use in classification and prognosis in school, but is of no convincing value as a measure of general intelligence as such, nor of demonstrated value in vocational guidance and direction.

IX. By JOSEPH PETERSON

George Peabody College for Teachers

1. The nature and measurement of intelligence. Intelligence seems to be a biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behavior. It is a mechanism for adjustment and control, and is operated by internal as well as by external stimuli. The degree of a person's intelligence increases with his range of receptivity to stimuli and the consistency of his organization of responses to them. In certain higher learning processes the bright subject more quickly and systematically enlarges the range of his cues or "guide-posts," and consequently more readily masters the situation than does the dull one, who must come back again and again to the more immediate elements. The latter cannot keep in mind the complexity of the situation. This difference is found to hold for children of very different years but equal mental ages, contrary to the results of a well-controlled experiment by Woodrow, one, however, the results of which seem to me to have been generalized too much by him and by some other writers. A dull and a bright child of equal mental age have gone through equal mental-age levels, but the bright child seems to have made the same acquirements in the lesser time.

One's emotionality, responsiveness to abstract symbols, degree of energy and of perseverance, etc., certainly condition one's intelligence. We cannot analyze such differences here, but can only suggest that equal mental ages by any test do not indicate just equal intelligence for all situations. We are probably too prone in our practical work of testing to assume equality of intelligence in such cases, since our attention is taken up with the one kind of reaction most susceptible to testing. General intelligence, if it is a reality at all, is probably not a separate or constant factor, but a composite of many different abilities, and probably means different things in unlike situations, as different abilities are stressed. Such factors as energy and perseverance, degree of disturbance by emotions and self-consciousness, and many others that play their roles in one's success in life, have not yet been successfully brought into the field of measurement by tests, especially by group tests. . . .

X. By L. L. THURSTONE

Carnegie Institute of Technology

What is meant by intelligence? Intelligence as judged in everyday life contains at least three psychologically differentiable components: a) the capacity to inhibit an instinctive adjustment, b) the capacity to redefine the inhibited instinctive adjustment in the light of imaginally experienced trial and error, c) the volitional capacity to realize the modified instinctive adjustment into overt behavior to the advantage of the individual as a social animal. . . .

These three phases of intelligent behavior may be labelled for convenience a) inhibitive capacity; b) analytical capacity, and c) perseverance. One may possess the inhibitive capacity without keen powers of analysis. This leads occasionally to an appearance of intelligence by silence. Such a mind is often capable of recognizing and adopting the successful analyses of other minds without being able itself to produce the analyses. The inhibitory capacity for the early, abstract, impersonal stage of instinctive adjustment does not necessarily coexist with inhibitory capacity for the later more defined and personally significant stage of the instinctive adjustment. Hence a man may possess inhibitory capacity for abstract impersonal thinking and lack self-control for those instinctive adjustments which have reached uninhibited a personally and socially significant and overtly defined stage. A man may possess considerable inhibitory and analytical capacity and still be socially quite ineffective because of insufficient volitional drive to realize his intelligence into consistent action. Such a mind is not infrequently the subject of ridicule from those who would justify their own mental worth in spite of modest analytical powers.

Analytical capacity concerns the preconscious and early conscious stage in the definition of instinctive adjustment. It implies the capacity to inhibit instinctive pressure at that stage so as to allow flexibility of definition of the response. This has been picturesquely called by Knowlson free interplay with the subconscious. It is only natural to expect that such inhibitory capacity should appear when the instinctive drives are relatively weak. Hence superior analytical ability is not infrequently associated with inferior instinctive energy in its nutritional, sexual, gregarious, and self-assertative forms. When such an individual discovers his inferior instinct equipment he may center his self-satisfaction on his analytical powers as a compensation. Such individuals need the assistance of other volitionally stronger minds in order to make their labors socially effective.

When the analytical capacity is associated in the same mind with a fair degree of instinctive and volitional energy we have the bright person in practical life. There is a fundamental distinction between bright and profound intelligence. The mind which possesses a fair amount of inhibitive, analytical, and volitional capacity will in general aim for effectiveness in the immediate present. Such a mind is bright. When, however, a mind is more generously endowed in all three of these respects or when it is unbalanced by reduced volitional capacity, it deals with the earlier and more abstract relations in its adjustments. Such a mind is profound and its analyses concern temporarily remote ends. I am quite sure that our intelligence tests in which the candidate races against time for a few minutes do not measure adequately the more inhibited and deliberative profound type of intelligence. It does fair justice to the more frequent well balanced bright mind. . . .

XI. By HERBERT WOODROW

University of Minnesota

1. The nature and measurement of intelligence. In attempting to state the meaning of intelligence, I should say first that it is an acquiring capacity. One's actual, present capacity for doing any particular thing is obviously dependent upon his training and learning, or, in general, upon his past experience; whereas, in the measurement of intelligence, the attempt is always made to minimize the effects of past experience by utilizing performances, the acquisition of the capacity for which is presumably favored equally by the past environments of all the individuals measured. When, for example, we use a child's present capacity to count backwards from twenty as a partial index of his intelligence, the assumption is, clearly, that the past environments of all children have been equally favorable for the acquisition of this ability. If this ability was one which could be readily taught, and, further, one which was taught some children but not taught others, it would be valueless as an index of intelligence. The present capacity to count is a measure of intelligence only in so far as it is an

S.C.E.R.T., West Bengal
Date...12.3.70...
151
S.C.E.R.T., West Bengal Library

index of the capacity to *learn* to count, that is, the capacity to acquire the capacity to count. Intelligence, then, is the capacity to acquire capacity. . . .

Now, if intelligence is to be defined as a capacity to acquire capacity, it is imperative to define the acquired capacity. I believe this capacity may be correctly described (using neither purely mental nor purely behavioristic terms) as the capacity for such mental activity as is most effective in bringing about success, or, in slightly different phraseology, as the capacity for success in so far as success is dependent upon mental processes, either past or present. By success, is here meant, not success in any one line of endeavor, but a theoretical success, namely, success in all those innumerable performances which may be regarded as desirable. The capacity for such a theoretical success would have as its practical result success in coping with any situation in which the individual might find himself.

This conception of intelligence implies that mind is an instrument for the securing of desirable ends. It implies that between mind and body there exists a relationship—a fact which I believe no living psychologist denies—but does not specify the ultimate nature of the relationship. Intelligence cannot be defined in purely mental terms, because the capacity for acquiring valuable modes of mental functioning is itself not mental; it is certainly, in part, a matter of the condition of the brain. Moreover, success can be attained only as the immediate result of behavior. All "mental measurements," even those concerned with "intensity of sensation," are primarily measurements or descriptions of behavior under measured or described environmental conditions. On the other hand, intelligence cannot be satisfactorily described in purely behavioristic terms, because (I mention only one reason, one which is persistently overlooked by behaviorists) while the degree to which behavior is intelligent is simply the degree to which it attains success, success has no real meaning except by reference to some want, desire, intention, plan, or purpose. . . .

XIII. By M. E. HAGGERTY

University of Minnesota

1. Intelligence and its measurement. In responding to this request to participate in the symposium on intelligence I take it that what is desired is a more or less dogmatic confession of faith rather than a systematic effort to support by evidence and argument the beliefs expressed.

1. In my thinking the word intelligence does not denote a single mental process capable of exact analytic definition. It is a practical concept of connoting a group of complex mental processes traditionally defined in systematic psychologies as sensation, perception, association, memory, imagination, discrimination, judgment and reasoning. The nearest equivalent to this working concept is James' "differentia of reasoning" as "the ability to deal with novel data" and Dewey's well known "forked-road" theory. This limitation of the concept of intelligence to the "novel-situation" category, so useful for theoretical purposes, however, seems to me too thin and lean for practical purposes of controlling behavior. I prefer to enrich the concept with all the fatness afforded by the inclusion of the more elementary and subsidiary processes instrumental to the crucial act of thinking or reasoning. For the most part, I would exclude from the concept, emotions, instincts, will-activities and so called character traits.

2. Intelligence is a dynamic concept. It is descriptive of behavior and not of static component parts of the "mind." When one conceives of the mind as a "state" or a "structure," the word intelligence becomes meaningless. The implication of *activity* is essential in the concept of intelligence.

3. Intelligence is native and peculiar to an individual. In this sense it is an instinct defined as an inherited or congenital characteristic. It may be conceived of as akin to, if not dependent on, or a phase of, the native quality of the phyical organism, the brain and nerves. Because of the original differences in intelligence or native capacity the same environment affects differently different persons. As a result two persons subjected to the same environment, say the formal procedures of the elementary school,

emerge with different mental content, different habits, different interests. These eventuating differences are the criteria by which we may infer the quality of the native endowment of the individual. Often these differences among individuals are less important in themselves than they are for such inferential purposes. Present attainments are important in showing what an individual can now do. In so far as such attainments may be made the basis of inference as to native endowments they are of value for prognosis. Because of their superior prognostic value intelligence tests have a value which mere achievement tests do not have.

4. Intelligence is not merely a qualitative concept. It may be quantitatively considered. When a child has little of it he does poorly in school, and in cases of extreme deficiency he is largely uneducable. When a child has more intelligence he is capable of learning and individuals gifted with great intelligence may learn most rapidly. Similarly in actual life low intelligence means simple occupations and crude civilization. High intelligence means the possibility of efficiency and leadership in the more complex problems of civilized society.

2

Intelligence as the Tests Test It

EDWIN G. BORING

The second selection is another early paper, very short and written primarily for nonpsychologists, but the distinction it makes between intelligence in general and intelligence as the tests test it has influenced much subsequent thinking. It is by Edwin G. Boring, "Intelligence as the Tests Test It," from The New Republic, *1923,* 34, *35–36, and is reprinted with the permission of the author.*

If you take one of the ready-made tests of intelligence and try it on a very large number of persons, you will find that they succeed with it in very different degrees. Repeat the test, and you will find that they cannot, with the best will in the world to do well, alter their scores very greatly. Then give the same group another intelligence test, and you will discover that the differences among individuals are approximately, although not exactly, the same. And you can go on. You will find that an adult, after continued exposure to his social and educational environment, does not greatly alter his score on a given test; that children, however, do steadily improve their performances until somewhere between ten and twenty years old; that the average age at which improvement stops is about fourteen years; but that children while improving tend to maintain the same individual differences, so that in a given group every child would keep about the same rank within the group. These are basic observational facts of the psychology of intelligence. What do they mean?

WHAT THE TESTS TEST

They mean in the first place that intelligence as a measurable capacity must at the start be defined as the capacity to do well in an intelligence test. Intelligence is what the tests test. This is a narrow definition, but it is the only point of departure for a rigorous discussion of the tests. It would be better if the psychologists could have used some other and more technical term, since the ordinary connotation of intelligence is much broader. The damage is done, however, and no harm need result if we but remember that measurable intelligence is simply what the tests of intelligence test, until further scientific observation allows us to extend the definition.

An observational method for extending knowledge of intelligence as the tests test it is the method of statistical correlation. The relation to intelligence of any measurable capacity at all can be determined by comparing the relative performances of a large number of persons in an intelligence test with their achievement in the measure of capacity in question. If the correlation is considerable, yet not perfect, say 60 per cent, we say that the particular capacity is partly dependent upon intelligence and partly independent of it. We shall not be far wrong if we think of such a capacity as complex, involving 60 per cent of intelligence and 40 per cent of some special ability that is not intelligence.

The method of correlation gives us at once some insight into the nature of intelligence as the tests test it. No satisfactory intelligence test exists at present which employs a single type of mental operation. Most tests for intelligence, like the army tests, consist of batteries of single tests, every one of which appears, on inspection, to test some special ability, like arithmetical ability, or an appreciation of verbal relations or of logical relations. When one obtains the correlations among the different tests that make up the battery called an intelligence test, one finds that the separate tests do not correlate with one another so very highly—not so highly as a rule as does one combined intelligence test with another. These results are explained by saying that the separate tests are really tests of separate abilities, and that each of these abilities involves, in part, intelligence, which is a factor common to all the tests, and in part a special ability, which is not intelligence and which therefore explains the failure of the tests to correlate very highly. When the separate tests are combined in a total score, the special abilities, being unrelated, are supposed to cancel out, leaving the score to represent the "common factor," intelligence.

Thus we see that there is no such thing as a test for pure intelligence. Intelligence is not demonstrable except in connection with some special ability. It would never have been thought of as a separate entity had it not seemed that very different mental abilities had something in common, a "common factor."

A CONFUSION OF MEANINGS

One of the most frequent reasons for the misunderstanding of the tests is the fact that the existence and importance of these special abilities are usually lost sight of. The psychologists themselves are very apt to forget them and it is no wonder that their lay audiences are scarcely aware of them. Yet it is not even possible to understand the nature of tested intelligence without considering them. They are forgotten in part because the "common factor" has seemed especially important and the interest of the testers in the last decade has centered in it. Words, however, have also helped to obscure their existence. The tested intelligence of an individual is often called his "mental age"; the increase of intelligence in childhood is generally called "mental growth." In this way psychologists have inadvertently equated the "intelligent" to the "mental," overlooking in their terminology the vast number of special abilities that help to make up the "mind." It is high time for a change of words here. The present usage requires us to say that the average adult has a "mental age" of about fourteen and that "mental growth" on the average stops at fourteen. Nothing could be more untrue. The statement can be true only of intelligence as the tests test it. The special abilities, which make up skill and knowledge, continue to cumulate presumably throughout all adult life.

A very useful conception of intelligence, and one that is approximately correct in the light of our present knowledge, is that intelligence is like "power" as the physicist uses the word: the amount of work that can be done in a given time. All intelligence tests involve the maintenance of time-limits to some extent, and most tests are "speed" tests where all the work is performed against time. We may think, then, of intelligence as power and of a special ability as a machine that utilizes the power for a particular purpose. No machine can operate without power, and power is actually demonstrable only when it has a machine through which to operate. It is idle to speculate as to which is the more important, the power or the adaptive device for the utilization of the power; and it is folly to bet one's fortune on the power, forgetting the machine.

Part II

INTELLIGENCE AND LEARNING ABILITY

The term “learning ability” is often loosely used as a synonym for “intelligence.” The fact that at all educational levels correlations between scores on intelligence tests and ratings of school achievement turn out to be fairly high seems to support such an interpretation. But a more careful analysis of the matter uncovers many complications. Both school records and intelligence-test scores are based on what the person has already learned *in the years he has lived. Thus the correlations may simply reflect the fact that questions asked in intelligence tests, like questions asked in school, give an advantage to subjects who for one reason or another have learned the most. To prove that such tests measure the ability to learn, it is necessary to answer the question: Do test scores* predict *the* amount *individuals will learn or the* improvement *in performance they will show under standard conditions?*

Besides pointing up an interesting theoretical issue, answers to this question have important practical implications. The uncritical assumption that intelligence tests measure general learning ability can lead to pessimistic prognoses for persons who score low and unrealistic expectations for persons who score high. Neither of these attitudes facilitates the learning every individual must accomplish if he is to contribute as much as he can to our complex society.

3

Intelligence and Improvement in School Subjects

HERBERT WOODROW

The first selection reports some significant studies by Herbert Woodrow, whose important work on the problem of learning ability has not always received the attention it deserves. This paper is an abridged version of "Intelligence and Improvement in School Subjects," from the Journal of Educational Psychology, *1945,* 36, *155–166. It is reprinted with the permission of the American Psychological Association.*

In view of the vast number of studies which have been made of the intercorrelations of achievement scores in a given grade, it is surprising that so few have been devoted to intercorrelations of measured improvements from one grade to the next. One reason for this neglect may be that the ability to improve has often been assumed to be synonymous with intelligence. (*1, 2, 3, 4, 5*) That little support for this view can be found in the gains made by grade-school pupils, over a period of a year or longer, is obvious from the data presented below. Two recently completed studies will be described. The results obtained suggest that prevailing concepts concerning the determining conditions of the gains made by school pupils are still in need of considerable development.

The measures of improvement used in each study were based on objective test-scores yielded by standardized achievement tests in the following six school skills or subject-matters: reading, vocabulary, arithmetic fundamentals, arithmetic problems, English, and spelling. The average published reliability coefficient of these part scores is .86.[1] The data have been studied particularly for their bearing on two questions: (1) whether the ability to improve is synonymous with intelligence, and (2) since the answer to the first question turns out to be negative, whether there is any general ability to improve, that is, any common factor, intelligence or otherwise, manifested in the gains in the various school subjects.

It should, of course, be kept in mind that the apparent determining condition of gains in school, that is, school attendance for a stated time in a given school class, although a more natural or life-like condition than practice conducted in a laboratory, is, in reality, a complex of variables which are not kept constant for all the pupils. Some variables which readily come to mind are irregularities in attendance at school, interest and liking for the subject, health, study habits, level of aspiration, various home influences, and a complex of conditions, not too well understood, determining what is rather vaguely called motivation. In spite of the lack of experimental control of conditions in the school situation, it is interesting to observe that gains in various school subject-matters or skills, at least in the upper grades, show about the same lack of relation with each other and with intelligence as do the gains observed under better controlled laboratory experiments on the effects of practice. (*6, 7, 8, 9*)

The first study, carried out in the grammar grades of the small town of Litchfield, Illinois, deals with improvements occurring in one calendar year, in the case of pupils of the fourth, fifth, and sixth grades. Since none of the children repeated grades in this school, upon the second testing the fourth-grade children were in the fifth grade, the fifth-grade children in the sixth grade, and the sixth-grade children in the seventh grade. The Metropolitan Achievement Battery was given to all pupils on the same half-day in May of successive years, different forms being used each year. . . . Each of the gain scores used consists of the difference between the two scores yielded by the Metropolitan Tests on successive years. The index of intelligence used was an average of scores obtained on two testings by the Otis Quick-Scoring

---

[1] Metropolitan Achievement Tests, Supervisor's Manual, World Book Co., 1939, p. 72 (Tests 1, 2, 3, 4, 5, and 9, grades IV–VIII incl.).

Mental Ability Test, Form Alpha one year, and Form Beta the following year.

Before giving the correlations between improvement scores, it may be well to refer briefly to the correlations of achievement scores. These correlations were as large as, or perhaps larger than, those usually found. A few were over .80, and most of them were between .50 and .70. The average of the fifteen intercorrelations between the six achievement scores at each testing are shown in Table 1. The mean of all the correlations represented in the table, numbering forty-five, is .60.

TABLE 1

Average Achievement Score Intercorrelation

Group	First testing	Second testing
5th grade	.536	.626
6th grade	.656	.641
7th grade	.557	.557

In spite of these very considerable correlations between achievement scores, the correlations between the gains occurring in one year were very low indeed. These correlations are given in Table 2. Included also are the correlations shown by the separate subject-matter gains with average IQ. Of the forty-five intercorrelations between gain scores, only three are large enough to be significant by the criterion of three times their standard error. Of these three significant correlations, two were correlations between gain in arithmetic problems and gain in arithmetic fundamentals, one of .468 in the fifth grade and one of .417 in the sixth grade. Nine of the correlations were negative, and the average of the entire forty-five correlations was only .12, a figure which contrasts strongly with the average intercorrelation of .60 obtained between the test scores for any one date and grade. Certainly the low average intercorrelation of .12 between the gains lends no encouragement to the idea that these gains are determined to any important extent by one and the same general ability to improve in mastery of school subjects during one year of study.

TABLE 2

Intercorrelation of One-year Gains in Achievement

	R	V	AP	AF	E	S	CA
V	.198						
AP	.208	.287				4th to 5th Grade ($N = 94$)	
AF	.148	.347	.468				
E	.126	.095	−.028	.071			
S	.212	.131	.168	.261	.005		
CA	−.267	−.321	−.213	−.250	.061	−.316	
IQ	.422	.341	.508	.432	−.065	.395	−.675
V	.112						
AP	−.119	.229				5th to 6th Grade ($N = 96$)	
AF	.043	.210	.417				
E	.115	.035	.052	−.039			
S	−.032	.192	.005	.224	−.106		
CA	−.202	−.257	−.180	−.140	−.009	−.092	
IQ	−.001	.233	.306	.273	−.127	.264	−.720
V	.183						
AP	.177	−.022				6th to 7th Grade ($N = 102$)	
AF	.156	.092	.193				
E	.156	−.024	.050	−.107			
S	.080	.216	.064	.134	−.112		
CA	−.162	−.073	−.062	−.015	−.165	−.178	
IQ	.169	.124	.076	.012	−.028	.250	−.754

The abbreviations represent the one-year gains in scores, as follows: R, reading; V, vocabulary; AP, arithmetic problems; AF, arithmetic fundamentals; E, English; S, spelling. CA stands for chronological age and IQ for the average IQ of the first and second testings.

The conclusion reached by direct examination of the intercorrelations of the gains was confirmed and extended by means of factor analyses carried out separately with the intercorrelations of each grade, by means of the centroid method, with rotation of axes. The rotation of the factors is admittedly a somewhat arbitrary one and does not pretend to reveal what Thurstone calls primary abilities. The results are shown in Table 3.

TABLE 3

Correlations of Gains with Rotated Orthogonal Factors

	I	II	III	IV	h^2
Fifth-grade Group					
R	.440	−.024	.278	−.029	.2723
V	.347	.374	.213	.084	.3127
AP	.433	.449	−.013	−.034	.3904
AF	.368	.546	.079	.288	.5227
E	−.068	.017	.435	.105	.2052
S	.433	.002	.050	.229	.2424
CA	−.707	−.072	−.001	.205	.5469
IQ	.933	.219	−.107	−.136	.9484
Sixth-grade Group					
R	.050	−.091	.421	.105	.1990
V	.130	.341	.199	.322	.2765
AP	.165	.541	.076	−.042	.3264
AF	.166	.683	.002	.028	.4949
E	−.107	.009	.321	−.080	.1209
S	.107	.187	−.132	.261	.1319
CA	−.856	.062	.334	.095	.8571
IQ	.800	.259	−.150	.324	.8346
Seventh-grade Group					
R	.151	.129	.482	.263	.3408
V	.121	.001	.059	.425	.1987
AP	.039	.304	.240	.001	.1515
AF	−.030	.489	.050	.153	.2659
E	.036	−.120	.381	−.099	.1707
S	.255	.177	−.100	.408	.2728
CA	−.936	−.028	.124	−.103	.9099
IQ	.825	.125	.004	.047	.6984

h^2, the sum of the squares of the four factor loadings, is the proportion of the total variance of a gain accounted for by the four common factors.

In the two higher grades, the conclusion that there was no general learning factor at all in the one-year gains is established to a high degree of certainty. In the lowest grade, that is, for the gains made from the fourth to the fifth grade, all gains but one showed positive correlations or "loadings," varying from .35 to .44, with the same factor. Only the gain in English showed no positive correlation with his factor (the correlation in this case being −.068). The lack of correlation of gains in English with this factor, in the fifth-grade group, is verified by the very low correlation of these gains with those in arithmetic fundamentals and arithmetic problems, .071 and −.028, respectively. Of course, it is possible that some special local conditions accounted for the low correlations shown by the English gains; and the fact that all other gains showed a positive correlation with the common factor, suggests the possibility that in the primary grades all gains in particular achievement tests might, as a rule, show a significant correlation with one general factor. This factor, which in the fifth-grade group was a general factor except for English, correlated .933 with IQ, and could, therefore, presumably be identified as largely the same thing as intelligence as measured by the Otis Tests.

In the two higher grades, also, there was revealed an intelligence factor, i.e., a factor which correlated highly with IQ. For grades VI and VII, the correlations of this factor with IQ were .800 and .825, respectively. In these grades, however, the correlations between the gain scores and this intelligence factor varied around zero (from .255 to −.107), and none of them can be considered statistically significant. For the sixth grade, they averaged .085; and for the seventh, .095. In these two grades, then, the gains may be said to show no general intelligence factor.

Besides the IQ factor, two, and perhaps three, other factors were found. One of these (factor II in Table 3) was a factor common to the two arithmetical gains, those in arithmetic fundamentals and in arithmetic problems. These two gains had a considerable loading with this factor in all grades, and in the fifth and sixth grades vocabulary gain also correlated with this factor to a degree possibly significant (.374 in Grade V and .341 in Grade VI).

A third factor (factor III in Table 3) had to do with gain in verbal performances, particularly reading and English. This factor showed correlation with gains in reading of .278, .421, and .482 in the fifth, sixth, and seventh grades, respectively. With gains in English, it showed correlations of .435, .321 and .381. The variation in the loadings of these verbal tests in the different grades may possibly be attributed to differences in course content and teaching. On the other hand, it may well be supposed that the nature of the factor itself varied somewhat with the correlational matrix from which it was extracted. In no grade, however, does this factor have any marked relation to Otis IQ; and it is, of course, since the extracted factors were uncorrelated, entirely unrelated to the chief factor in Otis IQ, that is, to factor I. In the case of the first two factors, it is quite certain that much the same factor is involved in all grades, the first of these two factors being identifiable by its very high correlations with IQ and the second by the fact that the two sorts of arithmetical gain are always definitely correlated with it. Only in the case of these two factors is it possible to postulate an approximate identity in the case of the different grades. A fourth factor was extracted, but its significance is slight and it appears very doubtful that it has the same interpretation in the case of the different grades.

Lastly, attention should be called to the fact that even when the four common factors extracted are all taken together, they account for only a part of the total variance in the gain scores. This is shown by the columns headed "h^2," the proportions listed in which may be interpreted as the sum of the variances produced by the four uncorrelated factors; and, further, if the total variance in any one gain is taken as 1.0, these proportions also represent the proportions of the total variances of the various gains accounted for by the joint effect of the common factors. Now these proportions averaged as follows: fifth-grade group, .324; sixth-grade group, .258; and seventh-grade group .233. The conclusion is indicated that by far the larger part of the variance in gains is due either to errors of measurement in the gains or to specific factors, that is, factors varying from gain to gain and not common to the gains in any two subject-matters. In all probability both errors of measurement (sometimes called errors of response), and specific factors are important causes of the gain variances. The relative importance of these two causes, however, cannot be specified because of lack of knowledge of the reliability of the gain scores. It is not possible even to state the reliability of the single scores, since it would be unsafe to assume this reliability to be the same as the published reliabilities (averaging .86). One might guess the average reliability of the single tests to be about .80, and the reliability of the gains to average somewhere between .40 and .50. If the gains have an average reliability of .40, which is certainly a very conservative estimate, it would still follow that specific factors were, roughly and on the average, about half as important (that is, contribute about half as much to the total variance of the gains) as all four of the group factors taken together.

In general, then, results of this study may be summarized as follows:

1) No gain factor was found in the case of any grade which had to do with gain in all six of the subject-matters.

2) Instead, the intercorrelations revealed three group factors: I. A factor which was exceedingly important in determining IQ score and which may be termed intelligence; II, a factor common to arithmetic fundamentals and arithmetic problems; and III, a verbal factor having to do with scores in reading and English. Factor I, the most important in determining Otis IQ, had practically nothing to do with gains in any of the school subjects in either the 6th or 7th grade.

3) For the most part, the gains as measured were unique, that is, due neither to any one of the common factors nor to all of them taken together, but to a combination of errors of measurement and the specific factors. It is probable that gain in each subject depended at least half as much upon its own specific factor as upon all the common factors taken together. . . .

[*The omitted portion deals with a briefer report of a study in Providence, Rhode Island, in which very similar results were obtained.*]

In general, it is obvious that neither the Litchfield nor the Providence data lend any encouragement to the identification of intelligence with learning ability as indicated by gain in school subjects. This statement is made with the qualification that learning ability is measured by the gain in score over some fixed period. If one takes as one's measure of learning ability the pupil's total achievement, that is, the difference of score at any given time from zero, then obviously the correlation between intelligence and ability to learn is identical with the correlation between intelligence and achievement score at a given time. This correlation is a pronounced one, particularly if corrected for attenuation. It is significant, however, that the correlation between intelligence and achievement reaches its maximum rather early in the school grades. This is not what one would expect if gains depended mainly on intelligence. One might then reasonably expect correlation with intelligence to show a marked increase in the higher grades as compared with the lower grades. While, then, the present data do not disprove the allegation that intelligence may be the chief determiner of how much the individual has learned in the past (measured from zero), they strongly indicate that intelligence cannot be postulated generally and without qualification to be the chief determiner of how much the pupil will learn in the future. If the first measurements of intelligence and achievement are made beyond the fifth grade, the indications are that intelligence, far from being the principal factor, is a rather minor factor in accounting for individual differences in improvement in any specified number of additional years. Moreover, in such a case there is no evidence of any other general factor, that is, a general unitary learning ability other than intelligence possessing an all-pervading importance as a determiner of future gains. Instead, amount of future gain seems (for reasons which the present study does not divulge) to correlate chiefly with quasi-general and specific factors, the former varying from one group of subject-matters to another and the latter from one subject-matter to any different one.

REFERENCES

1. CARROLL, R. P. "What is intelligence?" *Sch. & Soc.*, 1928, *28*, 792 (quotation from S. S. Colvin).
2. DE WEERDT, E. H. "A study of improvability of fifth-grade school children in certain mental functions." *J. educ. Psychol.*, 1927, *18*, 547, 557.
3. ELLIS, R. S. *The Psychology of Individual Differences,* New York: D. Appleton, 1928, p. 62.
4. MCGEOCH, J. A. *The Psychology of Human Learning,* New York, Longmans Green, 1942, pp. 13 and 251.

5. Spearman, C. *The Abilities of Man,* New York, Macmillan, 1927, p. 20 (quotation from Buckingham).
6. Woodrow, H. The relation between abilities and improvement with practice. *J. educ. Psychol.,* 1938, *29,* 215–230.
7. Woodrow, H. The effect of practice on test intercorrelations. *J. educ. Psychol.,* 1938, *29,* p. 572.
8. Woodrow, H. Factors in improvement with practice. *J. Psychol.,* 1939, *7,* p. 64, Table 3.
9. Woodrow, H. Interrelations of measures of learning. *J. Psychol.,* 1940, *10,* 49–73.

4

Intelligence Test Scores as Indicative of Ability to Learn

J. W. TILTON

Four years after the Woodrow report appeared, J. W. Tilton raised some serious questions about its conclusions, and presented evidence of his own that under some circumstances, which he specified, intelligence tests do indeed correlate with gains *in school achievement. The decisive factor seems to be the inclusion in the achievement tests of enough* difficult *items so that the bright pupil can demonstrate what he has gained. The Tilton paper, entitled "Intelligence Test Scores as Indicative of Ability to Learn," appeared in* Educational and Psychological Measurement, *1949, 9, 291–296. It is reprinted with the permission of the author and* Educational and Psychological Measurement.

INTRODUCTION

If one starts with the assumption that the most dependable measure of learning is secured by subtracting a measure of status before from one after the period of practice, study or instruction, he finds himself in an arresting situation with regard to the definition of intelligence as the ability to learn. This is so because measured changes in score tend to be negligibly correlated with intelligence-test scores (*2*). What then shall we say about the validity of the use of intelligence-test scores in schools, and elsewhere, as indices of the ability to learn? Must we conclude from this reported negligible correlation, in spite of a long period of use for that purpose, that intelligence-test scores are not indicative of the ability to learn? The position taken in this paper is not a categorical yes or no. The paper is not a full consideration of the topic. The emphasis in the paper is that the measures of learning which have provided the low correlation are not an adequate basis upon which to claim invalidation.

DEFINITION OF AN ADEQUATE CRITERION

What are the qualifications for measures of gain, if they are to constitute a criterion for the validation of intelligence-test scores as indices of the ability to learn?

First, the reliability of the measures of gain must be known, and should be as high as that for end-scores. This seems like a trite observation, but it seems to be needed. The tendency is to be satisfied with having used end-tests of acceptable reliability, and to let it go at that. To use reliable end-test scores is good, so far as it goes, but it is not enough. Even with the coefficient of reliability for the end-tests as high as .95, the reliability of the difference may be zero. The end-tests contribute only unreliability to the measures of differences. All you can say for reliable end-tests, is that the more reliable the end-tests, the less handicap there is to be overcome in the measurement of gains. How reliable the differences are depends chiefly upon what happens to change the pupils between testings. In other words, since reliability is the ratio of true variance to obtained variance, the reliability of gains is a positive function of the differentiation, which learning introduces, and a negative function of error in measurement.

The trouble in practice is that we do not wait for learning to introduce enough differentiation. We think we are doing well if we allow a whole school year to elapse between the initial and final tests. But even after a year, reliabilities of gains, for instance, on the parts of achievement-test batteries, are still low. Whatever the reliability is, it needs to be measured and reported. Without such information most measures of gain may be assumed to be very unreliable, and to show very low correlations with intelligence-test scores, or with any other independent measure, merely because of that unreliability.[1]

But for the measures of gain to be sufficiently reliable, or at least of known reliability, is not enough. A second requirement must be set up. If the measures of learning are to be used as a criterion, against which intelligence-test or other scores are to be validated, as measures of the ability to learn, something must be required of the learning situation. It must provide an opportunity for each to learn according to his ability.

Data with regard to spelling, reported by Carroll (*1*), illustrate a situation in which such an opportunity does not exist. At the beginning of the year pupils were tested as to their ability to spell the words to be taught during the year. Had they been tested again at the end of the year, the brighter pupils could not have made the larger gains, assuming 100 per cent mastery, because they knew 81 per cent of the words at the beginning of the year whereas the duller pupils could spell only 40 per cent of them. The bright could gain only from 81 to 100 from the situation, while the dull could gain from 40 to 100.

Gains obtained under such circumstances, where there are clear limitations of opportunity, are obviously not suitable as criteria of the ability to learn. It is quite possible that such limits are often present in the school situation; limited offering, because of the tendency to organize content into years and semesters, limited instructional help because of the pressure upon teachers to get about 95 per cent of the class above the required standard. At any rate, not all measures of gain are suitable to serve as criteria, even if satisfactory so far as reliability is concerned. Intelligence-test or other scores may not be invalidated as indices of the ability to learn in situations in which the possibility of validation is curtailed.

In the third place, and finally, the possibility of validation must be uncurtailed not only in the teaching or learning situation, but also in the testing situation. It is not enough that each has the opportunity to learn according to his ability, he must also have an opportunity for the full amount of that learning to be reflected in the measure of gain.

---

[1] Those not familiar with the substantial correlation which may be obtained between initial or final test and gain (*5*) merely because of the correlation of errors are probably led erroneously to infer reliability from such correlations.

Here again, as in the case of reliability, the tendency is to err by assuming that if it is a good test it will give good measures of gain. But if the test is a good initial test (ranging in difficulty from 0 to 100 through a central tendency of 50 per cent) it will not be a good final test and will yield poor measures of learning. If a good distribution of measures of learning is to be obtained, it will not come from a good initial test but only from a good test for measures of learning. It can be secured only after the use and repetition of a wastefully long test. It will be made, presumably, of items ranging from 0 to 100 through 50 per cent difficulty. Only in this case, the difficulty to be examined is not beginning difficulty nor end difficulty, but difficulty in learning. Item analysis for this purpose will not be based upon the per cent answering correctly, but upon the per cent learning to answer correctly. Such an examination of the measures of gain would not alone provide sufficient assurance of the adequacy of the measures, but it would reveal some of the grosser inadequacies.

These, then, are the qualifications for an adequate criterion against which intelligence or other scores may be compared, as indices of the ability to learn. For the measures of gain which have provided the published low correlations, the reliability has not been determined, it is not known to what extent the opportunity to learn was curtailed, and, if the opportunity to learn was adequate, it is not known whether the tests permitted good measures of the learning. If the validity of status scores, as indices of the ability to learn, is to be checked, data are needed, and they need to be chosen specifically for that purpose.

TWO ATTEMPTS TO SECURE AN ADEQUATE CRITERION

There were, in the writer's files, two sets of data which it was thought might yield satisfactory criterion-measures of learning.[2]

---

[2] The seventh-grade data have been described by D. C. Knowlton and J. W. Tilton (3). For the twelfth-grade data, the writer is indebted to Clarence H. Elliott.

Both were in the social studies area, one at the seventh-grade level and one at the twelfth. The data to be reported here for the seventh grade are for a sample. The sample is composed of all cases with a full set of scores from 6 of the 9 control sections in an experiment. For the whole grade in the single junior high school, (N = 515), the mean Otis Group IQ was 100.0, the sigma was 22.2. For the sample (N = 134), the mean and sigma were 102.9 and 23.1. The sample is therefore fairly representative of the whole seventh grade. The data to be reported for the twelfth grade are for the full twelfth grade in one high school, N = 156.

The tests in the seventh grade were made for four[3] experimental units of instruction. They were composed of 300 items, and covered eleven weeks of instruction. The twelfth-grade test was the *Columbia Research Bureau American History Test*, composed of 200 items. The period covered was one school year.

In each case most of the items were discarded. What was wanted, as stated above, was to be sure of differentiation as to learning with the least curtailment of opportunity. The choices were arbitrary. First, it was assumed that when fewer students answered an item correctly at the end of the period than at the beginning, the loss was a chance difference. Then all items were discarded unless the gain was larger than the largest loss. Second, it was assumed that if the abler students were being penalized in the full test, it was by the presence of items answered correctly at the beginning by a majority of the students. Arbitrarily, for the seventh grade, items were discarded which initially were answered correctly by more than 45 per cent. For the twelfth grade, for which a greater caution seemed advisable, all items were discarded if answered correctly by more than 25 per cent at the beginning of the year.

There remained for criterion tests of learning, 54 items for the seventh-grade sample, and 20 items for the twelfth grade. The 54 items ranged in learning difficulty from 7 to 83 per cent, averaging 34 per cent. The 20 items ranged from 18 to 58 per cent, with an average of 38 per cent. The reliability coefficients are .93 and .73.

---

[3] Units 1, 3, 4, and 5. Unit 2 was discarded for the present purpose because of the relatively small gains made (3) Table 4, p. 42.

THE USE OF CRITERION SCORES

How well should intelligence-test scores agree with criterion measures of learning? What "passing mark" may be set, above which validity may be said to have been established? It is easy enough to agree that if the correlation is zero, validity has not been established. But is it reasonable to require the correlation to approach unity? There are two reasons why that would be unreasonable. One is that the factors of interest and application are unaccounted for. The other is specialization of ability. It has not been assumed that intelligence-test scores are equally indicative of learning in all areas. A reasonable requirement is that the agreement between index and criterion should approximate that between the index and initial status in the area of the criterion learning.

This requirement is met in the seventh-grade sample and in the twelfth grade. The correlations for the seventh-grade Otis IQ are with initial .50 and with the learning criterion .49. For the twelfth-grade Terman Group IQ's, the correlations are with initial .43 and with the learning criterion .49.[4]

DISCUSSION

As it turned out, in these two cases the construction of criterion measures with selected items was in a sense a wasted effort. The correlations of IQ with gains using all items were not reliably less than with the criterion measures. In other words, the gains on all items were proved to be good criterion measures, as they were—fairly reliable (.72 and .68), and obtained in a situation in which there was plenty of opportunity both to learn and to be measured. The criterion measures gave evidence of superiority only with respect to reliability, .93 for 54 items as compared with .72 for all 300 items, and .73 for 20 items, as compared with .68 for all of the 200 items.

---

[4] Correction for attenuation does not alter the picture much. In this correction the reliability of the Otis IQ was estimated, the other reliabilities were computed. The four corrected correlations in the order mentioned above are: .58, .53, .56, and .58.

CONCLUSION

It may be said that within the two sets of data described, intelligence-test scores predicted learning fairly well. But this paper has been presented, not so much for the weight of the evidence presented, as to serve as a stimulus to similar reports, a plea for better data than are available. In hundreds of experiments, intelligence-test scores have been used in equating groups. The correlation between these scores and gains should be reported, with measures (or good estimates) of the extent to which the gains are suitable criteria. If intelligence-test scores are to be discredited as indices of the ability to learn, it must be done with more concern for defensible criteria than has been given to the problem.

REFERENCES

1. Carroll, H. A. *Generalization of Bright and Dull Children.* New York: Teachers College, Columbia University Bureau of Publications, 1930.
2. Findley, W. G., Turnbull, W. W. and Conrad, H. S. "Construction, Evaluation, and Applications for Intelligence Tests." *Review of Educational Research,* XVII (1947), 14–15, 17.
3. Knowlton, D. C. and Tilton, J. W. *Motion Pictures in History Teaching.* New Haven: Yale University Press, 1929.
4. Simrall, Dorothy. "Intelligence and the Ability to Learn." *Journal of Psychology,* XXXIII (1947), 27–43.
5. Zieve, L. "Note on the Correlation of Initial Scores with Gains." *Journal of Educational Psychology,* XXXI (1940), 391–394.

5

Studies in Concept Attainment: Effect of Instruction at Two Levels of Intelligence

S. F. OSLER AND S. R. WEISS

Tilton's results pointed up the fact that the relationship between intelligence and learning ability depends in some way on the difficulty of the material to be learned. On achievement tests that included a fair proportion of difficult items, gains were correlated with intelligence ratings. More recent studies have begun to analyze in finer detail the qualitative differences in the kinds of learning that occur in brighter and duller children. In a series of experimental investigations of the attainment of concepts, Sonia F. Osler and various associates have produced evidence pointing to several such differences. The report included in this collection, by S. F. Osler and S. R. Weiss, is entitled "Studies in Concept Attainment: III. Effect of Instructions at Two Levels of Intelligence," Journal of Experimental Psychology, *1962, 63, pp. 528–533. It is reprinted with the permission of the author and the American Psychological Association.*

In a recent study on the effect of intelligence on concept attainment using instructions which were nonspecific as to the nature of the task, Osler and Fivel (1961) reported a significant difference in errors between groups of average and superior IQ. An important question that has yet to be answered concerns the pertinent attribute of high intelligence which accounts for the superiority of performance. One possibility is that superior *Ss* supplement *E*'s instructions with their own, directing them to search for consistencies in the reinforced stimuli; whereas the less intelligent *Ss* work along without self-instructions until the reinforcement contingencies of the experiment strengthen the response to the concept exemplars. Perhaps a simpler statement of this hypothesis might be that under the conditions of nonspecific instructions the task presented to *S* has two components, namely, problem *finding* as well as problem *solving*; and that the superior performance associated with high intelligence may be the result of greater proficiency at either or both of these components. On the other hand, when instructions are specific, defining the problem which *S* is expected to solve, the problem finding component of the task is eliminated or reduced to a minimum.

It was the aim of the present investigation to test these ideas experimentally. Two experiments were performed. Experiment I was a replication of the Osler and Fivel (1961) experiment. This was undertaken for the double purpose of confirming the earlier data and controlling for the possibility that automated equipment presently being used for the conduct of the experiment may have affected the results obtained. In Exp. II the general instructions were replaced by more specific ones, defining the problem for *S* and directing him to search for hypotheses.

METHOD

As Exp. I and II differed only in the instructions used, the method for both will be described together, and wherever necessary information pertinent to each experiment will be provided separately.

Subjects

The *S*s consisted of 180 school children attending elementary and junior high schools of Baltimore City and County, equally divided among the following age groups: 6, 10, and 14 yr. Each *S* was within 6 mo. of the age specified. Half of the *S*s within each age level had a WISC IQ between 90 and 109 (Average Intelligence group), and the other half had a WISC IQ of 110 and above (High Intelligence group). The mean IQs of the two groups in Exp. I were 100.8 and 120.9, and in Exp. II they were 101.2 and 121.7 respectively.

Apparatus

The apparatus has been described elsewhere (Osler & Powell, 1960). Essentially, it consisted of a film strip projector and a screen to receive the projected images of the stimuli which appeared in pairs. A telegraph key placed under each stimulus released a marble following a correct response, which consisted of pressing the key under the concept exemplar. No reinforcement followed if the key under the negative stimulus was pressed. The *S*s were not paced, but 4 sec. after a response was made the next stimulus was presented.

The concept studied was *bird*. This was one of the three concepts on which data were reported by Osler and Fivel (1961). The stimuli used were the same as in the previous investigation, except that they now appeared on a screen projected from a film strip. The complete set of stimuli consisted of 150 pairs of pictures, one member of each pair representing a positive instance of the concept and other a negative instance. The pictures of each pair were matched approximately in area, color, and brightness. Reasons for using naturalistic stimuli, which carry with them the hazard of unspecified dimensions, have been discussed elsewhere (Osler & Fivel, 1961). An additional reason prompting their use in this investigation was the previous finding that performance on this concept was a function of intelligence, thus demonstrating the suitability of this problem for the present study. The stimuli had been pretested on a group of children of the same age and intelligence levels as the present Ss, and were found to be familiar to the type of subject studied.

Procedure

Each *S* was tested individually, in two separate sessions. At the first session *S* was told that he would play a game in which he could win any one of a group of toys displayed before him. He was asked to select the most preferred of the group, and this was then set slightly apart from the others in a position which allowed

S to view it during the "game." The *S*s in Exp. I received the following instructions:

> This is the game we are going to play. Here are two handles (*E* points to them). If you push the correct handle you will get a marble. Sometimes this will be the correct handle and sometimes this will be the correct handle. If you push the correct handle you will get a marble like this (demonstrated). If you push the wrong handle, like this, no marble will come (demonstrated). If you get a marble you can put it here (here *E* points to a grooved wooden tray with which *S* was provided). Now you try this one and let's see what happens. If you look at these pictures before pushing the handle, you will be able to get many marbles. When you fill the tray with marbles, you can have the (name of the toy).

The stimuli used during the demonstrations were unrelated to the concept.

In Exp. II the *S*s were also informed about a game in which they could win a toy, and then they received the following instructions:

> This is the game we're going to play. Listen carefully and I will tell you how to play the game. Here are two pictures. You pick one of them and push the handle under it, and if you are right, you will get a marble, like this (demonstrated). If you are wrong you will not get a marble, like this (demonstrated). Now you try and push this one. The game is to see how soon you can get a marble every time you choose. If you look at the picture carefully you will see that there is something in the pictures like an idea that will tell you which one to choose to get a marble every time. Now we will start the real game. What are you going to do first? (if *S* does not say: look at the pictures, *E* supplies the answer). If you look carefully and see which pictures give you the marble you will be able to win marbles every time. Are you ready? Remember, look at the pictures first. (The preceding sentence was repeated together with the following sentence whenever necessary.) Remember, there is an idea that will tell you which picture will give you the marble.

The procedure was continued until *S* made 10 consecutively correct responses, which was the criterion of success. At this point *E* commented on *S*'s good work and asked him how he knew where the marble was each time. If *S* could not state the rule correctly the procedure was continued with questions following each block of 10 consecutively correct responses until *S* gave a correct answer or completed 150 trials. Any *S* who did not achieve criterion performance at that point was considered to have failed. A second session, held within 2 weeks of the first, was devoted to the administration of the WISC.

Experimental Design

In both experiments a 3 × 2 factorial design was used with three ages and two intelligence levels constituting the respective variables. Thus there were six experimental groups with 15 *S*s in each group, a total of 90 *S*s in each experiment.

RESULTS

The results were analyzed in terms of three measures of performance: (a) the mean number of errors to criterion, (b) the number of *S*s who attained the criterion, and (c) the number of *S*s who verbalized the concept at the point of success.

Experiment I

The mean number of errors per group and the results of an analysis of variance,[1] presented in Table 1, indicate that age and intelligence were associated with reliable differences in performance, thus reproducing the findings reported by Osler and Fivel (1961). Similar results were obtained from an analysis of the data on frequency of *S*s who attained criterion performance and those who were able to verbalize the concept, as shown

---

[1] Among the 14-yr-old *S*s the conditions necessary for application of an analysis of variance were not entirely satisfied because of skewness of the distribution and the smaller variance of the 14-yr-old High Intelligence group. For this age group the distribution of errors was examined and compared (by X^2) for average and high intelligence within and between the two experiments. There was no evidence of significant differences among the 14-yr-old *S*s according to intelligence or between the experiments.

TABLE 1

Mean Number of Errors to a Criterion of 10 Consecutive Correct Responses with General Instructions

	Groups						
	6-Yr. *S*s		10-Yr. *S*s		14-Yr. *S*s		
Intelligence	Mean	*SD*	Mean	*SD*	Mean	*SD*	Means by IQ
Average IQ	71.2	21.0	38.9	34.5	14.3	22.2	41.5
High IQ	57.1	27.8	19.7	17.0	6.5	11.4	27.8
Means by Age	64.2		29.3		10.4		

Note.—$F_{\text{age}} = 37.66$, $df = 2$, $P < .001$; $F_{\text{intel.}} = 7.14$, $df = 1$, $P < .01$; $F_{\text{Age} \times \text{Intel.}} = <1$.

in Table 2. In summary, then, it turned out that with instructions that did not specify the nature of the task, age and intelligence were associated with improved achievement in concept attainment on each of three measures of performance.

Experiment II

Table 3 presents the mean number of errors per group and the results of an analysis of variance of errors, respectively, for Exp. II. In this experiment, in which the nature of the task was explicitly defined, the effect of intelligence disappeared; only age was associated with significant differences in the number of errors. Similar results were obtained from an analysis of the other two measures of performance, number of *S*s who achieved crite-

TABLE 2

Number of S*s* *Who Reached Criterion and Number Who Verbalized the Concept under General Instructions*

	Groups							
	6-Yr. *S*s		10-Yr. *S*s		14-Yr. *S*s		All Ages	
Intelligence	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.
Average IQ	4	1	8	2	13	13	25	16
High IQ	7	4	14	11	15	14	36	29
Total	11	5	22	13	28	27	61	45

Note.—Criterion data: Age $\chi^2 = 22.690$, $df = 2$, $P < .001$; Intelligence $\chi^2 = 6.156$, $df = 1$, $P < .02$. Verbalization data: Age $\chi^2 = 33.068$, $df = 2$, $P < .001$; Intelligence $\chi^2 = 7.512$, $df = 1$, $P < .01$.

TABLE 3

Mean Number of Errors per Group for Reaching Criterion of 10 Consecutive Correct Responses with Explicit Instructions

	Groups						
	6-Yr. *S*s		10-Yr. *S*s		14-Yr. *S*s		
Intelligence	Mean	*SD*	Mean	*SD*	Mean	*SD*	Means by IQ
Average IQ	46.1	25.5	34.2	26.5	14.1	17.4	31.5
High IQ	46.5	34.0	37.0	31.3	7.7	17.4	30.4
Means by Age	46.3		35.6		10.9		

Note. — $F_{\text{age}} = 13.52$, $df = 2$, $P < .001$; $F_{\text{intel.}} < 1$; $F_{\text{Age} \times \text{Intel.}} < 1$.

rion performance and number of *S*s who verbalized the concept, as shown in Table 4.

Combined Data from the Two Experiments

In view of the comparable variances obtained in both experiments the error data were combined and analyzed for a possible Instructions × *S* interaction. The results of this analysis, presented in Table 5, show that while instructions produced no main effect on performance, the Age × Instruction interaction was significant at the .05 level. When the data were subsequently analyzed separately for each age group, it turned out that only for the 6-yr. *S*s did instructions produce a significant difference in performance ($F = 5.92$, $df = 1.59$, $P < .05$) in favor of explicit-

TABLE 4

Number of Ss *Who Reached Criterion and Number Who Verbalized the Concept under Explicit Instructions*

	Groups							
	6-Yr. *S*s		10-Yr. *S*s		14-Yr. *S*s		All Ages	
Intelligence	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.
Average IQ	11	6	12	8	14	12	37	26
High IQ	9	4	9	9	14	14	32	27
Total	20	10	21	17	28	26	69	53

Note.—Criterion data: Age $\chi^2 = 7.081$, $df = 2$, $P < .05$; Intelligence $\chi^2 = 1.552$, $df = 1$, *ns*. Verbalization data: Age $\chi^2 = 17.715$, $df = 2$, $P < .001$; Intelligence $\chi^2 = .046$, $df = 1$, *ns*.

TABLE 5

Analysis of Variance of Errors for Exp. I and II Combined

Source	*df*	*MS*	*F*
Instructions (I)	1	616.05	
Age (A)	2	29,819.62	45.05**
Intelligence (IQ)	1	2,456.81	3.71
I × A	2	2,382.52	3.60*
I × IQ	1	1,798.67	2.72
A × IQ	2	7.90	
I × A × IQ	2	407.87	
Within	168	661.89	
Total	179		

* $P < .05$.
** $P < .001$.

ness. Another finding, rather surprising, and one difficult to explain, was that the bright 10-yr. *S*s were *handicapped* by the explicit instructions. The mean number of errors for these *S*s under the two conditions was 19.7 and 37.0, respectively ($t = 2.57$, $df = 28$, $P < .02$). The oldest group, the 14-yr. *S*s, performed identically under both conditions. However, as is apparent from their scores, these *S*s performed near the maximum possible level, and, therefore, could not be expected to show improvement under any conditions.

The analysis of the combined frequency data of successful *S*s shows that aside from age, reliable differences were produced by instructions in the Average Intelligence but not in the High Intelligence group. The same results were yielded by the data on correct verbalization.

Rank-order correlations were computed between error scores and MAs in each experiment. The rho values obtained were −.68 under general instructions and −.51 under explicit instructions. Both values are significant at the .001 level. Since the MA is heavily determined by the CA, the value of the correlation per se does not throw light upon the role of the IQ alone. However, as *S*s in both experiments were identical in age and intelligence, the difference between the two values of rho may be attributed to variations in experimental conditions. Unfortunately, no statistical test exists for evaluating the difference between the rho values; therefore all that can be said is that the direction of the difference is consistent with the other findings.

DISCUSSION

Under general instructions high intelligence was associated with superior achievement in all three measures of performance; but this superiority was not maintained under explicit instructions. If the assumption that specific instructions eliminate the problem-finding component from the task is correct, then the conclusion to be drawn is that high intelligence gave *S* an advantage in problem definition but not in problem solution in the concept attainment task. It is entirely possible, however, that were the concept task one of greater difficulty, intelligence might also prove to be advantageous in the solution phase of the task. This requires further investigation.

Since age was associated with more effective concept attainment on all three measures of performance and under both types of instructions, it may be concluded that maturation involves growth in problem *solving* ability at least until some point between 10 and 14 yr.

In order to assess changes in problem *finding* ability as a function of age, it is necessary to compare performance at each age level under the two types of instructions. That 6-yr. *S*s improved under the explicit instructions is apparent from the significant Age × Instruction interaction in Table 5. The situation was not clear for the 10-yr. *S*s because they actually deteriorated under explicit instructions. As this finding seemed to be inexplicable on any logical basis and was inconsistent with the results pertaining to the other two age groups, it was decided to replicate this portion of Exp. II. Upon replication it turned out that the average 10-yr. *S*s profited from the explicit instructions, while the superior *S*s were unaffected. As to the 14-yr. *S*s, whose performance was the same under both conditions, they did not have much room for improvement, precluding the possibility of reaching a definite conclusion on the effect of instructions at this age.

It thus appears that at least between 6 and 10 yr. *S*s benefited from explicitness of instructions. It is therefore reasonable to conclude that the more effective performance associated with increasing age is not brought about by improvement in problem finding, but rather by more effective problem solving. This points to a possible interesting difference between mental age and intelligence, the effectiveness of the latter being due to more effective problem finding rather than problem solving.

SUMMARY

A concept attainment task was presented under general and explicit instructions to *S*s 6, 10, and 14 yr. of age, divided into two intelligence groups. The results were analyzed in terms of three measures: errors, number of *S*s who achieved the criterion of success, and number of *S*s who verbalized the concept correctly.

Under nonspecific instructions superior intelligence was associated with more effective concept attainment on three measures of performance, but under explicit instructions *S*s of average intelligence improved while those of superior intelligence remained unchanged. Apparently, superior intelligence gave *S*s an advantage in the problem finding phase of the task, but not in actual problem solution.

Performance improved throughout the age range studied. There was no evidence for improved problem finding ability between ages 6 and 10 yr., and it was difficult to evaluate the data for the 14-yr.-olds. It appears, therefore, that problem solving ability, but probably not problem finding ability, increases within the age range studied.

REFERENCES

Osler, S. F., and Fivel, M. W. Concept attainment: I. The role of age and intelligence in concept attainment by induction. *J. Exp. Psychol.*, 1961, 62, 1–8.

Osler, S. F., and Powell, M. G. Apparatus for the study of discrimination and concept formation. *Amer. J. Psychol.*, 1960, 73, 627–629.

Part III

INTELLIGENCE AND TEMPERAMENT

6

Cognitive, Conative, and Non-Intellective Intelligence

DAVID WECHSLER

One of the issues that has divided theorists since the early days of intelligence testing is whether or not the cognitive and non-cognitive aspects of individual differences can really be separated. Everyone agrees that both kinds of factors enter into every achievement, in school and in life. But we have a long historical tradition of distinguishing between two aspects of human nature, such as reason and emotion, ability and motivation, ego and id. To many psychologists it has seemed important that this distinction be clearly maintained in the construction of tests to evaluate human characteristics. Others have held that the distinction is an artificial one, impossible to apply in practice, and that intelligence testers should frankly admit that motivation and personality characteristics are an ineradicable feature of every "mental measurement."

The progress of research has tended to support those who take this latter position. Like the "learning ability" issue, this one also has pervasive practical implications. Especially in schools, whose function it is to stimulate intellectual growth, it makes a considerable difference whether we think of wishes and attitudes as component parts of intelligence or as things completely independent of it.

David Wechsler, whose individual tests for adults and children are essential tools used by all psychologists concerned with the assessment of intellectual level, has insisted throughout the years that intelligence and temperament cannot be divorced. The paper reprinted here is a comprehensive statement of his position, "Cognitive, Conative, and Non-Intellective Intelligence," Ameri-

can Psychologist, *1950, 5, pp. 78–83. It is reprinted with the permission of the author and the American Psychological Association.*

It is always a good omen for science when different men in different places make independent discoveries or arrive at similar conclusions. In the last two decades psychologists in their efforts to define the nature of general intelligence seem to have arrived at the threshold of such a situation. In this paper I wish to present to you what appears to me to be the germ of the impending reorientation: it is this, that general intelligence cannot be equated with intellectual ability however broadly defined, but must be regarded as a manifestation of the personality as a whole.

From an historical point of view, the first one to argue against the identification of general intelligence with intellectual ability was Henri Bergson. Already in his "Donées Immédiate de la Conscience" and more emphatically in his "Evolùtion Créatrice," he pointed out the insufficiencies of the human intellect or, what was for him the same, normative logic, in dealing effectively with man's total environment.

I shall not here restate Bergson's arguments nor his attempted solution of endowing the human mind with a new faculty, creative intuition, and its generating force, the "elan vital." I wish only to call your attention to the fact that in our attempts at measuring intelligence we have persisted in treating intelligence as if it consisted exclusively of intellectual elements or factors. What, in fact, are these intellectual elements which we have continued to use and to posit in appraising intelligence? They are abstract reasoning, verbal, spatial, numerical, and a few other specified factors, all of which in some particularized manner deal with man's cognitive ability. Shades of Bergson, are we confirming his claim that human intelligence, as the psychologist conceives it, can only deal with geometric and logical symbols?

Now, the remarkable thing is that while this is what we are saying in our tests of intelligence, most of us don't believe it. What is more important, it isn't true! Our contemporary definitions of intelligence assert as much: intelligence according to these is not only the ability to learn, to abstract, to profit from experience, but also to adjust and to achieve. Everyone with clinical experience knows that the latter involve other capacities besides eductive, verbal, numerical, spatial, and the other intellective factors that have been demonstrated. Yes, but what are they? The answer is: they are *not* intellective. They are capacities and traits dependent upon temperament and personality which are not restricted to logical and abstract perception; they are, in my opinion, factors of personality itself. It is this point of view, independently sensed or suggested, at times only tangentially, by a number of investigators including Goldstein, Alexander, Wechsler, and more recently by Halstead and Eysenck, which I presented six years ago for the first time under the term *Non-intellective Factors of Intelligence*. I wish now to present to you more fully the evidence in its support and to justify what appears to be not only the need for a re-orientation in our concept of general intelligence, but of a new psychometric that will, in fact, measure what is purported in our definition of intelligence.

Let me begin by restating the issue in terms of the actual psychometric problem. The crux of this problem, as we have already noted, is the discrepancy between what the clinical psychologist does and what he says he does in clinical practice. If we examine any of the current psychological tests of intelligence, we shall find them to consist of sample tasks measuring, by definition, a variety of mental abilities. One would imagine that any summary of the results obtained with such tests would be essentially a report of the degree to which an individual possesses these abilities and the manner in which they vary. However, it will be found that once a summative score is obtained from them, whether in terms of MA, IQ, or whatnot, the clinical psychologist proceeds to enlarge his summary to include not only specific psychologic interpretations but broad social and biological implications as well.

An IQ is thus used, not only to determine comparative mental endowment, capacity to learn, presence of special abilities and disabilities, and evaluation of degree of mental deficiency, but also as a basis for school placement, for vocational guidance, for psychiatric diagnosis, and for the prediction of adjustment potentials in a variety of situations from infancy to old age, including such areas as child adoption, juvenile delinquency, fit-ness for military service, college success, and old age counseling.

Assuming that intelligence tests may be used in all these situations, and within limits I believe they may, the question arises how this is possible under the concept that general intelligence is a matter of a single basic or even a combination of a number of intellectual abilities. It is this question which I shall try to answer this evening. But I must first call your attention to the fact you are all aware of, that this is not the usual criticism of intelligence tests. The historic and continued objection to intelligence tests is not that they measure too much, but that they do not measure enough, or at least, not well enough.

You are all acquainted with the arguments against intelligence tests, and I shall not repeat them; the damaging criticism pertains, not as is generally emphasized, to the question of reliability, but to one of basic validity. Even such studies as those of Wellman, Goldfarb and others, showing changes in IQ produced by a variety of social and environmental factors, though relevant, are not crucial. The crucial instances are those where individuals obtain identical IQs (say an IQ of 65) but, on overall appraisal, must nevertheless be rated differently, say, one as a defective and the other as not defective. Such instances are not necessarily common, but neither are they rare exceptions. Here is a situation which needs explaining and cannot be by-passed.

The first to attack this problem was E. L. Thorndike. His answer, as always characteristic of his approach, was straightforward and to the point. Our tests measure intelligence to be sure, he said, but there is not just one unique, but several different kinds of intelligence, namely, abstract, social and practical. The first is manifested by the individual's ability to work with symbols, the second by his ability to deal with people, and the third by his ability to manipulate objects. Thorndike, himself, seems to have been primarily interested in the first kind of intelligence and, having made the above trichotomy, and along with it the distinction between tests which measure breadth, as against those which measure altitude, left the working out of these concepts to others. But relatively little has been done to verify or refute the hypothesis.

In the 1920's Moss published a test of social intelligence which consisted essentially of items involving memory and recognition of names and faces, and a series of multiple-choice questions involving social situations, in which the correct answer seemed to have been based on the notion that "the customer is always right." Although Moss's test for a time had some vogue among business firms, clinical psychologists, as far as I have been able to discover, seldom if ever make use of it.

The other important effort at producing a test of social intelligence is Doll's Vineland Social Maturity Scale. This Scale, as you know, consists of a series of questions listing a variety of social acquisitions, that is, of approved and useful acts and achievements, which a child may be expected to have learned from infancy to adolescence. The Scale is hardly a test in the ordinary sense of the term, since it involves no test performance or response by the subject, and can be completed, as it usually is, by other persons. But it does correlate fairly well with other tests of intelligence and has been shown by Doll and others to correlate positively and significantly with a number of practical criteria of social adjustment.

Clinical psychologists appear to have accepted performance tests, almost from the start, as a measure of practical intelligence. Only they seem to have regarded practical intelligence, as measured by these tests, as a kind of special aptitude rather than as a kind of intelligence. For many years the situation in clinical practice was something like this: a child would be given routinely a Binet test. Then, if his Binet MA did not seem to do justice to him, he would be given a Pintner-Paterson or similar performance battery as a supplementary test. But the child's score on the performance test, except in instances of language handicaps, would seldom be integrated with, or serve to alter, his Binet intelligence rating. Instead, it would usually be used as evidence of a compensatory useful special ability. Thus, if a child attained an IQ of 85 on the Binet, and one of 110 on the Pintner-Paterson, the reporting psychologist would ordinarily give the rating as "dull normal" intelligence with good practical or manipulative ability. It was not until the publication of the Bellevue Scales that any consistent attempt was made to integrate performance and verbal tests into a single measure of intelligence. The Bellevue tests have had increasingly wider use, but I regret to report that their popularity seems to derive, not from the fact that they make possible a single global rating, but because they enable the examiner to obtain separate verbal and performance IQ's with one test.

The Aristotelian hierarchical white-collar concept of intelligence dies hard. This, in spite of the fact, that performance tests often can and do contain a larger amount of *g* than do the verbal tests. Thus, in his differential study of "Abstract and Concrete Intelligence," W. P. Alexander, after correcting for communality, specific factors, and chance errors of measurement, found the theoretical *g* loadings for verbal and practical ability to be .60 and .81, respectively. Alexander concludes that "a perfect performance battery would be a better measure of *g* than a perfect verbal battery."

This and other findings by Alexander bring me to what constitutes the most compelling evidence for the reorientation in our concept of intelligence mentioned at the onset of this paper. I refer to the findings contributed by factor analysis. Here two important names appear on the horizon: Carl Spearman and L. L. Thurstone. I believe that the answers which they have given to the problem of the nature of general intelligence are incorrect. But I am sure that without the inspiration and without the tools which they furnished us, the solution of the problem would be altogether impossible.

Such a statement before a gathering of clinical psychologists may be unorthodox, because to many, factor analysis is almost anathema. But I can assure you, on the authority of expert consultants, that the mathematics of factor analysis is quite elementary, and on the basis of my own experience with it, extremely practical; and, with due apologies to Freud, even "sexy." For with what, in effect, does factor analysis concern itself, but with the bedfellowship of psychometric tests. For, mind you, it embraces matrices, correlational to be sure, and then tells you what test stays close to what other tests when axes are rotated. Now that, I submit, is what clinical psychologists want to know: what test, what factor, or, if you will, what function or what trait goes with what other factor, or function, or trait. And when the findings are examined some very interesting and unsuspected relationships come to light. For example, some tests of intelligence, like some human beings, are extremely promiscuous. Thus, vocabulary, the paragon of verbal tests, correlates very frequently, and to a considerable measure, with Block Designs, the perfect example of a performance test. But to return to a more serious vein, the importance of factor analysis is, of course, that it enables us to discover what our tests measure and the extent to which they measure the things they purport.

What are the elements which factor analysis has shown our intelligence tests to measure? The first is abstract reasoning. This is Spearman's *g* or eduction. Spearman argued that *g* was the only independent factor, and while he hesitated to identify *g* with general intelligence, his actual applications are tantamount to it. In equating *g* with general intelligence Spearman was in error, not because the tetrad equation is incorrect but because in point of fact, it is not satisfied as he claimed. Spearman's answer to this finding was that we cannot expect the tetrad equation to be satisfied by all the tests of general intelligence but only by "good" tests of intelligence, like analogies and mathematical reasoning which require eduction. But of course, if you select your tests, you can choose them so highly saturated with a single factor that the residuals vanish. This is all that the tetrad equation says, and it was the perceptive insight of Thurstone which recognized the tetrad equation for what it was, namely, a mathematically special case of a more general solution of the factorial problem. What was needed was a statistical analysis which would permit the emergence of other factors when present. By the use of his expanded technique, it has now been shown that intelligence tests, such as they are, contain not one but several independent factors. Some five or six have been definitely identified; they are, to repeat, induction, verbal, spatial, numerical, and one or two other factors. Notice, however, that these factors, like Spearman's eduction, are all cognitive.

At this point it is important to bear in mind what a factor stands for in factor analysis. Basically, it is an identifiable independent variable which accounts for a certain portion of the total test variance in a correlational matrix. The amount of variance it accounts for in any given test is called the test's factor loading. In a perfectly factorialized correlation matrix, the sum of the factorial loadings of the extracted factors should be 100 per cent, that is, account for the total test variance.

Now, it is a remarkable finding that when matrices of intelligence tests are factored, the amount of variance accounted for is seldom more than 60 per cent of the total, and, what is perhaps of equal significance, the greater the number of different tests included, the smaller, generally, the total per cent of variance accounted for; and this is seemingly independent of the number of factors extracted. In the case of our present intelligence test batteries, factors beyond the first 3 or 4 usually contribute so little to the already accounted-for variance that it is generally not profitable to extract them. It is the observation of this important finding that in the factorialization of batteries of intelligence tests, there always remained a considerable per cent of unaccounted-for variance, which began to arose my interest some years ago. It seemed to hold the key to our problem.

If after successive attempts at factoring out all the components of intelligence, there always remained a large residue of these unknown elements, the obvious inference to be made was that our intelligence tests measured other things than those accounted for by the extracted factors. The second inference was that those other factors were numerous and occurred in relatively small amounts, because it was impossible to extract single additional factors which would account for any considerable portion of the residual variance. I assumed that the principal reason for this was that the test batteries usually factored did not include tests which contained sufficient amounts of these other factors, to enable some of the remaining tests to cluster about them. Provisionally I called these residual components the nonintellective factors of intelligence. But in terms of more recent findings, I believe they can be more justly designated as the personality components of general intelligence, which in fact they are.

The evidence for this conclusion comes from a number of sources. As early as 1913, Webb (*8*), in factoring a battery of tests, along with a number of ratings which attempted to appraise traits of character, was able to extract a factor "W." "W" in a broad sense seemed to relate to a moral and conative propensity, which he called conscientiousness or purposeful consistency. A few years later, in Spearman's own laboratory, Lankes and Wynn Jones (*7*) demonstrated the existence of another non-intellective factor, "p," or perseveration, which characterized their subjects, tendency to resist changes in set, and which Spearman related to his law of inertia. In 1921, W. M. Brown (*2*) discussed character traits as factors in intelligence tests, and in 1933, R. B. Cattell (*3*) reported correlations between tests of temperament and ratings in intelligence. But perhaps the most crucial findings are those of W. P. Alexander (*1*) who, in an extensive factor analysis of a large series of verbal and performance tests, supplemented by tests of achievement and academic marks, showed that in addition to the now familiar *g*, V (verbal ability), and P (practical ability), a considerable portion of variance had to be ascribed to two other extracted factors, namely, X and Z. X was a factor which determined the individual's interests and "concerns," in Alexander's words, "temperament rather than ability"; while Z was "an aspect of temperament related to achievement," in the case of Alexander's subjects, to school achievement.

The factor loadings of X and Z varied greatly from test to test, but even some of Spearman's ostensibly pure tests of *g* contained some Z and nearly all the performance tests showed considerable X or Z loadings. As might be expected, these factors played an even greater role in academic or technical achievement. For success in science, for example, the X factor loading was .74, as against only .36 for *g*, and for English .48 as against .43 for the *g* loading. From these findings one might even infer that lack of intellectual ability, beyond a certain point, accounts for relatively little of school failures. Indeed Dorothea McCarthy (*6*) recently offered the "hypothesis," and I quote, "that emotional insecurity . . . is the basic cause of most educational disabilities and learning failures, which are not due to mental defect."

What are we to make of these two findings? First, that factors other than intellectual contribute to achievement in areas where, as in the case of learning, intellectual factors have until recently been considered uniquely determinate, and, second, that these other factors have to do with functions and abilities hitherto considered traits of personality. Among those partially identified so far are factors relating primarily to the conative functions like drive, persistence, will, and perseveration, or in some instances, to aspects of temperament that pertain to interests and achievement. This, to be sure, is just the beginning, but one of the reasons that not much more has been done is that psychologists have continued to assume that personality has little to do with intelligence. To Thurstone as well as to Spearman, general intelligence seems to be first and foremost a cognitive function, by Spearman to be accounted for by a single pervasive factor, by Thurstone by a number of factors.

It is curious that the clinical psychologist, so little impressed by, or, at least, so little conversant with factor analysis, has almost from the start dealt with intelligence test findings as if the personality components in intelligence were already an established fact. For what does psychological diagnosis on the basis of intelligence test findings consist of but inferring adjustive capacities of the subject as a persona? It appears that the clinician, like the character in Moliere's *Malade Imaginaire*, has been speaking prose all his life without knowing it.

One might add that diagnosing personality and personality disorder, at the level on which it is being done, is not very difficult. Practically every good individual test of intelligence lends itself to such application to a greater or lesser degree, the Bellevue Scales and the new Children's Test of Intelligence perhaps a little more readily. This does not mean that they are tests of personality, but they do suggest that our intelligence tests contain elements which are essentially factors of the personality as a whole rather than of specific cognitive abilities. When the neurotic does poorly on the Digit Span Test, it is not because of defective memory, but generally because of a basic anxiety mobilized by the test, as by any other situation, in which he is seemingly on trial. Conversely, when a mental defective does relatively well on the Maze Test, it is generally not because he has better planning ability, but because he is less impulsive. Similarly, a large variety of traits and personality factors may be inferred from test performance—for example, energy level from a subject's performance on the Digit Symbol, asocial tendencies from general comprehension, masculinity-femininity from the picture completion test. These are only a few of the traits and diagnostic constellations with which every clinician who has done psychological diagnosis is familiar.

The point here is not that personality traits can be discovered in psychometric performance, or, what needs no special argument, that personality and abnormal conditions influence intelligence test findings, but that personality traits *enter into* the effectiveness of intelligent behavior, and, hence, into any global concept of intelligence itself. It is one thing if a child does poorly on an intelligence test because he is disinterested or upset and quite another if he is congenitally impulsive or emotionally unstable.

One would naturally suppose that if intelligence is a function of the personality as a whole, one should find significant positive or negative correlations with measures of personality itself. Such, indeed, are the findings, but the results are extremely hard to evaluate. This is in part due to the fact that the studies in this area have been done primarily with the intent of discovering the extent to which intelligence accounts for variance in personality. In an article which appeared in 1940, Irving Lorge reviews the studies published to that date on the general relationship between measures of intelligence and various measures or estimates of personality. The personality tests included most of the current and older inventories (Woodworth, Laird, Thurstone, Bernreuter, Allport, et al.), as well as the association experiment and the personality measures of Hartshorn, May, and Maller. Some 200 correlation coefficients were analyzed. The range of coefficients was from +.70 to −.49 with a median of +.04. Disregarding the signs, half of the ratios were between .00 and .15, and one quarter of them .30 and above. Lorge's general feeling about the findings is that the range is so "extraordinary that anybody can make any statement." Nevertheless, his conclusion is "that some correlation between intelligence and personality exists" (*5*).

All this is rather meager fare, but the findings are perhaps as satisfactory as could be expected. Apart from the known unreliability of paper-and-pencil inventories, there is the more disturbing fact of their uncertain validity and relevance. At times they do not measure the traits claimed for them, at others they measure only small segments of the personality, although in different ways; and at still other times, traits which are purely nominal. The latter, for example, was shown by Flanagan (*4*) to be the case with the Bernreuter Inventory dichotomies. In the original publication the test was scored for six different traits, which by factorialization were then reduced to two.

Flanagan's study is a good example of how factor analysis aids us in getting at basic components. Mere evidence of concomitant variation is not enough; in fact, it is often misleading. For example, defective hearing may have a measurable effect on both learning arithmetic and size of vocabulary, but, obviously, has no basic relation to either arithmetical reasoning or verbal ability. A variable to be basic and scientifically significant must be independent. In the case of man's cognitive functions, these independent variables, in so far as they are relevant to general intelligence, have been pretty well identified. It may be possible to add one or two to Thurstone's list, but not many more. Those of personality are yet to be discovered. We have some knowledge of what the factors to be measured are likely to be, some on the basis of researches like those of Webb, Alexander, Guilford, Cattell, and Eysenck, others on the basis of general observation and clinical experience. The latter have thus far gone unrecognized, not only because we have no tests for them but because clinicians, like their more academic colleagues, still think of intelligence as consisting primarily of cognitive abilities. Any bit of behavior that seems concerned with or related to instinct, impulse, or temperament is ipso facto considered as having no direct relation to general intelligence.

Such, for example, is curiosity: This was one of the traits which Terman in his studies of genius found most frequently among his gifted children. But he did not have, nor do we as yet have, any test of curiosity. No attempt has been made to extract curiosity as a factor of intelligence. We all know how important curiosity is for biologic adaptation as well as scientific achievement. It is, to quote McDougall, "at the basis of many of man's most splendid achievements, for rooted in it are his speculative and scientific tendencies," and ". . . in men in whom curiosity is innately strong, it may become the main source of intellectual energy and effort." But what is curiosity? "It is the impulse to approach and examine more closely the object which attracts it," that is an instinct, and according to McDougall, one of the basic instincts.

One need not be afraid or ashamed to acknowledge impulse, instinct and temperament as basic factors in general intelligence.

It is indeed because I believe they are that I have brought before you the arguments and evidence presented. My main point has been that general intelligence cannot be equated with intellectual ability, but must be regarded as a manifestation of the personality as a whole. I have tried to show that factors other than intellectual enter into our concept of general intelligence, and that in everyday practice, we make use of them knowingly or not.

What is needed is that these factors be rigorously appraised. Factor analysis has been emphasized because, at present, it is the only method which enables us to demonstrate and discover independent variables. We already have some clues as to what the non-intellective but relevant factors of intelligence may be. What we now need are tests which not only identify but measure them. This in effect demands broadening our concept of general intelligence and calls for a revised psychometric to measure these added variables as sub-tests of all general intelligence scales.

To say that general intelligence can be social and practical, as well as abstract, was just a beginning. We had to know what basic components of the mind were responsible for making an individual effective in one rather than in another area.

To realize that general intelligence is the function of the personality as a whole and is determined by emotion and conative factors is also just a beginning. We now need to know what non-intellective factors are relevant and to what degree. This is the task which lies immediately before us.

REFERENCES

1. Alexander, W. P. Intelligence, concrete and abstract. *Brit. J. Psychol. Monogr. Suppl.*, 1935, 6, No. 19.
2. Brown, W. M. Character traits as factors in intelligence test performance. *Arch. Psychol.*, 1923.
3. Cattell, R. B. Temperament tests. I. Temperament. *Brit. J. Psychol.*, 1933, 23, 308–29.
4. Flanagan, J. C. Factor analysis in the study of personality. Stanford Univ. Press, 1935.
5. Lorge, I. *Intelligence: its nature and nurture*. 39th Yearbook. National Society for Study of Education, 1940, Part I, 275–81.

6. McCarthy, Dorothea. Personality and learning. *Amer. Counc. Educ. Studies,* 1948, Series I, No. 35.
7. Spearman, C. *The abilities of man.* New York: Macmillan, 1927.
8. Webb, E. Character and intelligence. *Brit. J. Psychol. Monogr. Suppl.,* 1915, III.

7

A Nonintellectual Intelligence Test

HARRISON GOUGH

Wechsler presented evidence that personality qualities are revealed by the kinds of answers people give to the sorts of questions intelligence testers ask them. Harrison Gough, approaching the matter from the opposite direction, has accumulated considerable evidence that intelligence can be assessed by the kinds of questions personality testers typically ask. The paper is entitled "A Nonintellectual Intelligence Test" and is reprinted from Journal of Consulting Psychology, *1953,* 17, *242–246. It is reprinted with the permission of the author and the American Psychological Association.*

The title of this paper may be something of a misnomer, for it is not at all certain that the testing instrument which is going to be presented is, actually, a measure of intellect. Nevertheless, it does correlate significantly with conventional tests of intelligence, and seems to reveal other properties which suggest that something along the line of intellectual ability, or perhaps more precisely intellectual efficiency, is being appraised. The other part of the title—the "nonintellectual" part—has not been the cause of any misgivings, because, as will be seen shortly, the test instrument does not call for any problem solving, reasoning, abstract thinking, or any of the other mental functions which are traditionally regarded as the key factors in defining and measuring intelligence.

The two immediate aims of this project were first, to determine whether a criterion-specific set of personality inventory items could be assembled which would correlate significantly with accepted measures of intellect, and second, if this was successful, to determine whether the individual items could be combined in a scale which would have anything approaching practical significance and value.

The second objective is added, because, as we know for instance from work on the Rorschach (*1*), a series of variables can easily be identified which will correlate significantly with intellect, but which will seldom rise above the level of +.20 and +.25 into a region of more dependable utility when pooled in a composite.

PROCEDURES

The first step in the present project was to write a series of criterion-specific inventory items, and also to select appropriate items from standard sources. An example of a specially written item would be, "I get sort of annoyed with writers who go out of their way to use strange and unusual words," and of an item from a standard questionnaire, in this case the MMPI (*5*), "I gossip a little at times."

These items were then reproduced in a booklet and administered to the senior students in four Minnesota high schools.[1] In each school, the responses of students in the upper 25 per cent of the range on a conventional intelligence test were compared with those of students in the lower quartile.[2]

Table 1 provides an example of the way in which the analysis was conducted. Items which showed consistent and significant differences in all Hi-Lo comparisons were retained for the scale. Altogether, 76 items were identified which met this criterion.

These 76 items were scored on the original samples with the results indicated in Table 2. The next step was to cross-validate this scale. Table 3 presents the results observed in six cross-

---

[1] The four communities were: Bemidji, Dawson, Redwood Falls, and St. Cloud. The writer wishes to thank the following persons for making this testing possible: Mr. R. B. Clay, Mr. C. H. Dahlin, Mr. H. B. Gough, Mr. R. Gray, Mr. J. W. Smith, Mr. R. H. Sorenson, Mr. G. Ramseth, and Mr. R. H. Witt.

[2] Because of the small number of cases available, schools A and B were combined into a single sample for this analysis.

TABLE 1

Item Statistics for Four Items from the Ie *Scale*

	Percent saying "True"					
	Higher scores on intellectual ability measures			Lower scores on intellectual ability measures		
Item	School A (*N* 28)	Schools B and C (*N* 21)	School D (*N* 48)	School A (*N* 28)	Schools B and C (*N* 21)	School D (*N* 48)
1. It is more important that a father be kind than that he be successful.	79	95	92	71	76	69
2. Parents are much too easy on their children nowadays.	36	43	46	50	71	67
3. I am not afraid of picking up a disease or germs from door knobs.	86	81	85	79	67	67
4. Most people make friends because friends are likely to be useful to them.	21	19	29	46	71	77

TABLE 2

Summary Statistics for the First (76-item) Intellectual Efficiency (Ie) *Scale in the Original Samples*

Sample	*N*	*M*	*SD*	*r**
School A	135	52.44	8.51	.47
School B	35	51.34	9.20	.68
School C	66	51.41	9.41	.64
School D	205	51.93	9.51	.54

* Intercorrelations between the *Ie* scale and a conventional IQ test in each of the schools.

validating samples.[3] In the first five, only MMPI (5) items were available, but the sixth sample took the full 76-item scale. All of the correlations are clearly significant, statistically, and all save one are high enough to be of some practical value. An interesting finding here was the degree to which the scale retained its validity when checked in the graduate student sample where one would encounter rather severe attenuation of the criterion measure.[4]

---

[3] The first four samples were made available through the kindness of Mr. M. J. Lunemann, Dr. Starke Hathaway, Mr. F. J. Herda, and Mr. H. B. Gough. The graduate student samples were seen at the University of California Institute of Personality Assessment and Research, Dr. D. W. MacKinnon, Director, as part of a large-scale study of personal and social effectiveness. The Institute projects are being carried out under a grant from the Rockefeller Foundation.

[4] The preliminary 59-item MMPI version was tried in the Michigan clinical psychology assessment program (*6*). The abbreviated *Ie* scale yielded significant correlations with the following criterion variables: diagnostic competence (.21); individual psychotherapy (.32); professional interpersonal relations (.28); over-all clinical competence (.23); preference for hiring (.24); and liking (.23).

The correlations for professional interpersonal relations and for individual psychotherapy were both the second highest obtained among all the objective scales utilized in the Michigan study (including the Miller Analogies Test, 6 Primary Mental Abilities test scales, 14 Guilford-Martin scales, 10 conventional MMPI scales, 59 scales on the Strong Vocational Interest Blank, 9 Kuder Preference Inventory scales, and 6 scales from the Allport-Vernon Study of Values).

The individual items were also checked in the two graduate student samples, and it was discovered that 24 of them were contributing only slightly, if at all, to the correlations with the criterion tests. It was decided, accordingly, to eliminate them and to reduce the scale to 52 items.[5] This 52-item scale is the one which is currently being used, and it is also the version in-

TABLE 3

Summary Statistics for the First Ie *Scale in Six Cross-validating Samples**

Sample	*N*	*M*	*SD*	*r*†
I. For the abbreviated (59-item) scale				
1. St. Cloud ninth grade	82	38.09	7.46	.53
2. Minneapolis ninth grade	190	39.44	8.24	.36
3. St. Cloud twelfth grade, 1947	231	39.55	7.50	.52
4. St. Cloud twelfth grade, 1948	271	40.30	7.70	.47
5. University of California advanced graduate students	40	48.58	4.02	.42
II. For the full (76-item) scale				
1. University of California first-year graduate students in psychology	30	63.20	5.12	.44

* The first five samples were scored only for the 59 items taken from the MMPI. The psychology sample took the full scale.

† For the two graduate student samples, the criterion test was the Miller Analogies Test, form g. For the other samples, conventional IQ measures were used.

---

[5] A table listing each item, the percentage in each sample responding "True," and the significance tests has been deposited with the American Documentation Institute. Order document No. 3972 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting $1.25 for 35-mm. microfilm or $1.25 for 6- by 8-in. photocopies.

cluded in the California Psychological Inventory.[6,7] These 52 items, and the scored responses, are listed below.

1. I have had very peculiar and strange experiences. (False) 2. It often seems that my life has no meaning. (False) 3. When someone does me a wrong, I feel I should pay him back if I can, just for the principle of the thing. (False) 4. Any job is all right with me, so long as it pays well. (False) 5. People often expect too much of me. (False) 6. I get sort of annoyed with writers who go out of their way to use strange and unusual words. (False) 7. Success is a matter of will power. (False)

8. When in a group of people I have trouble thinking of the right things to talk about. (False) 9. It is more important that a father be kind than that he be successful. (True) 10. I was a slow learner in school. (False) 11. I must admit I have no great desire to learn new things. (False) 12. My skin seems to be unusually sensitive to touch. (False) 13. Parents are much too easy on their children nowadays. (False) 14. I read at least ten books a year. (True)

15. I have the wanderlust and am never happy unless I am roaming or traveling about. (False) 16. My parents have often disapproved of my friends. (False) 17. Teachers often expect too much work from the students. (False) (18). I am not afraid of picking up a disease or germs from door knobs. (True) 19. I am quite a fast reader. (True) 20. I have had more than my share of things to worry about. (False) 21. I feel like giving up quickly when things go wrong. (False)

22. In a group of people I would not be embarrassed to be called upon to start a discussion or give an opinion about something I know

---

[6] The California Psychological Inventory, copyright 1951. This inventory is designed to assess a number of social and interactional variables. Its emphasis is not on medical, psychopathological, or maladaptive characteristics, but, rather, on more favorable and positive aspects of individual functioning. It contains 19 scales, including measures for social responsibility, flexibility, academic motivation, tolerance, etc. See, for example, references (3) and (4).

[7] Thirty-nine of the 52 items were taken from the MMPI (by permission of the copyright owners), and the remaining 13 items were original. The 39 MMPI items, and their scorings, are as follows (numbers given are for the MMPI group form booklet): 13F, 28F, 33F, 35F, 36T, 37T, 60T, 62F, 63T, 78T, 116F, 122T, 146F, 149F, 194F, 198F, 221T, 224F, 225T, 256F, 260F, 267F, 277T, 280F, 304F, 338F, 392F, 448F, 455F, 462T, 464T, 487F, 492F, 521T, 524T, 526F, 541F, 546T, 552T, 559F.

well. (True) 23. I like to read about history. (True) 24. In school I found it very hard to talk before the class. (False) 25. I dread the thought of an earthquake. (False) 26. I like science. (True) 27. I am bothered by people outside, on street cars, in stores, etc. watching me. (False) 28. I have never been in trouble because of my sex behavior. (True)

29. I often get feelings like crawling, burning, tingling, or "going to sleep" in different parts of my body. (False) 30. I seem to be about as capable and smart as most others around me. (True) 31. I like poetry. (True) 32. I am quite often not in on the gossip and talk of the group I belong to. (False) 33. A windstorm terrifies me. (False) 34. I enjoy a race or game better when I bet on it. (False) 35. Most people make friends because friends are likely to be useful to them. (False)

36. I have never seen a vision. (True) 37. The future seems hopeless to me. (False) 38. I have had no difficulty starting or holding my urine. (True) 39. I often feel as if the world was just passing me by. (False) 40. I get pretty discouraged with the law when a smart lawyer gets a criminal free. (True) 41. The only interesting part of the newspaper is the "funnies." (False) 42. I gossip a little at times. (True)

43. At times I have been so entertained by the cleverness of a crook that I have hoped he would get by with it. (True) 44. I have had attacks in which I could not control my movements or speech, but in which I knew what was going on around me. (False) 45. I have had no difficulty in starting or holding my bowel movement. (True) 46. I like to read about science. (True) 47. I do not read every editorial in the newspaper every day. (True) 48. If people had not had it in for me, I would have been much more successful. (False) 49. I daydream very little. (False)

50. I have often been frightened in the middle of the night. (False) 51. I work under a great deal of tension. (False) 52. I seldom worry about my health. (True)

A review of these items is quite interesting in highlighting some of the personological correlates of general intelligence. The best method of extracting the immediate implications of the items is simply to read the list carefully, but as an aid in forming rational groupings and clusters the following synopsis is provided:

1. Self-confidence and self-assurance, freedom from unsubstantiated fears and apprehensions.

2. Effective social technique and adjustment, sense of social acceptability without dependence on others; not suspicious, touchy, or overly sensitive.
3. Good physiological functioning, absence of minor, debilitating symptoms and complaints.
4. Liking, and respect, for intellectual pursuits; wide range of interests.

Three additional samples were then obtained in order to cross-validate the 52-item revised scale.[8] Table 4 presents the findings.

TABLE 4

Summary Statistics for the Revised (52-item) Ie *Scale in Three Cross-validating Samples*

Sample	*N*	*M*	*SD*	*r**
1. Rock Island, Illinois high school	461	33.90	7.07	.49
2. Wellsville, New York, high school	230	34.08	6.39	.50
3. University of California undergraduate psychology students	43	40.05	5.16	.42

* Intercorrelation between the *Ie* scale and a conventional measure of intelligence in each sample.

The validity coefficients are about what they were with the first six cross-validating samples, so it appears that the shortening did not weaken the scale's effectiveness.

Final high school grade averages were also available for the Rock Island sample. For the females, the correlation between grades and tested IQ was .56, and between grades and the *Ie* scale .59. For males the correlation between grades and IQ was .51 and between grades and the *Ie* scale .31.

In order to provide an over-all picture of the summary statistics at this point, the means and standard deviations for each

---

[8] The Rock Island, Illinois, sample was obtained through the courtesy of Mr. Charles O. Austin, Jr.; the Wellsville sample through the kindness of Dr. H. D. Babcock; and the University sample through the courtesy of Dr. David Krech and Mrs. Ann Miller.

TABLE 5

Ie *Scale Summary Statistics for All Samples**

Samples	*N*	*M*	*SD*	*r*	*SE*
I. Original samples					
1. Bemidji twelfth grade	135	35.87	5.82	.47	.09
2. Dawson twelfth grade	35	35.12	6.29	.68	.17
3. Redwood Falls twelfth grade	66	35.16	6.44	.64	.12
4. St. Cloud twelfth grade, 1949	205	35.52	6.50	.54	.07
II. Cross-validating samples					
1. St. Cloud ninth grade	82	31.80	6.52	.53	.11
2. Minneapolis ninth grade	190	34.74	7.26	.36	.07
3. Rock Island high school	461	33.90	7.07	.49	.05
4. St. Cloud twelfth grade, 1947	231	34.84	6.61	.52	.07
5. St. Cloud twelfth grade, 1948	271	35.50	6.78	.47	.06
6. Wellsville high school	230	34.08	6.39	.50	.07
7. University of California undergraduate psychology students	43	40.05	5.16	.42	.15
8. University of California first-year graduate students in psychology	30	43.23	3.50	.44	.19
9. University of California advanced graduate students	40	42.80	3.54	.42	.16

* Scores for samples not taking the final 52-item scale were proportionately corrected to equate them to this standard.

sample have been proportionately equated to the 52-item test version with the results indicated in Table 5. In this table one can observe a consistent rise in mean scores from junior high school groups, through senior high school and college undergraduate levels, to a final culmination in the graduate student samples.

ADJECTIVE CHECK LIST

The last data to be presented on this scale have to do with the "social-stimulus values" of persons scoring high and low on it. Forty graduate students seen in an intensive 3-day assessment at the University of California Institute of Personality Assessment and Research were given the scale. Staff judges completed Gough Adjective Check Lists (2) for each assessee, and then these observer-reports were composited. Adjectives differentiating significantly between the composite reports on the 10 highest and 10 lowest scorers on the *Ie* scale are given below:

1. Adjectives checked more frequently about higher scorers.
    - adaptable
    - appreciative
    - capable
    - clear-thinking
    - conscientious
    - considerate
    - cooperative
    - dependable
    - energetic
    - enterprising
    - helpful
    - honest
    - industrious
    - interests wide
    - kind
    - logical
    - obliging
    - persistent
    - planful
    - pleasant
    - reasonable
    - realistic
    - reliable
    - responsible
    - sincere
    - stable
    - thorough
2. Adjectives checked more frequently about lower scorers.
    - aloof
    - cynical
    - defensive
    - deliberate
    - effeminate
    - headstrong
    - impatient
    - individualistic
    - moody
    - resentful
    - sarcastic
    - sharp-witted

The general trend of adjectives ascribed is quite clear. High scorers are described as capable, clear-thinking, dependable, logical, realistic, etc., and low scorers as cynical, headstrong, impatient, and resentful.

The adjective findings lend support to the notion that the *Ie* scale is measuring something like personal or intellectual efficiency, with their emphasis on characteristics such as industriousness, planfulness, reliability, and thoroughness.

SUMMARY

A study was made of the relationship between intellectual measures and a series of opinion, attitude, and self-description items believed, on a priori and theoretical ground, to covary with intellect. Fifty-two items were found to possess empirical validity. These items were assembled in a scale which yielded median validity coefficents of approximately .58 in four original samples, and of .47 in nine check samples.

Adjectival descriptions of subjects scoring high on the scale tended to stress factors of dependability, intellectual clarity, persistence, and planfulness. The decision was made to call the scale a measure of "intellectual efficiency" in order to reflect these properties.

REFERENCES

1. ALTUS, W. D., and THOMPSON, GRACE. The Rorschach as a measure of intelligence. *J. consult. Psychol.*, 1949, 13, 341–347.
2. GOUGH, H. G. *Predicting success in graduate training*. A progress report. Berkeley: University of California Institute of Personality Assessment and Research, 1950.
3. GOUGH, H. G. Identifying psychological feminity. *Educ. psychol. Measmt.*, 1952, 12, 427–439.
4. GOUGH, H. G. What determines the academic achievement of high school students? *J. educ. Res.*, 1953, 46, 321–331.
5. HATHAWAY, S. R., and MCKINLEY, J. C. *The Minnesota Multiphasic Personality Inventory*. Minneapolis: Univer. of Minnesota Press, 1943.
6. KELLY, E. L., and FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. of Michigan Press, 1951.

Part IV

INTELLIGENCE OR INTELLIGENCES?

8

The Abilities of Man

CHARLES SPEARMAN

The question of whether "intelligence" is essentially a singular or a plural word has challenged psychologists since the scientific study of intellect began. The raw material for all the mathematical investigations of this problem is a table of correlations between scores a group of subjects make on several tests. These correlations always turn out to be positive; they never turn out to be perfect. What can we make of this combination of facts? The method by which meaning is extracted from large sets of intercorrelations is called factor analysis. *It has already been referred to in some of the previous papers in this book. (Wechsler's description of the method for clinicians constitutes a lighthearted introduction for nonmathematical readers.)*

The first serious studies of the organization of mental abilities were undertaken by Charles Spearman. The presentation reprinted here is from pages 73–82 of his book, The Abilities of Man *(New York: Macmillan, 1927). Footnotes and references to the Appendix have been deleted.*

Criterion of "tetrad differences." The start of the whole inquiry was a curious observation made in the correlations calculated between the measurements of different abilities (scores for tests, marks for school subjects, or estimates made on general impression). These correlations were noticed to tend towards a peculiar arrangement, which could be expressed in a definite mathematical formula. And the formula thus originally reached has ever since been maintained without any essential change. Only from time to time, for convenience, it has been converted from one form to some other that is mathematically equivalent.

The form recently preferred is given below. In it, as usual, the letter r stands for any correlation, whilst its two subscripts indicate the two abilities (tests, school marks, etc.) that are correlated.

$$r_{ap} \times r_{bq} - r_{aq} \times r_{bp} = 0.$$

This formula has been termed the *tetrad equation* and the value constituting the left side of it the *tetrad difference.*

An illustration may be afforded by the following imaginary correlations between mental tests (actually observed correlations will be given in abundance later on):

		Opposites	Completion	Memory	Discrimination	Cancellation
Opposites	1		.80	.60	.30	.30
Completion	2	.80		.48	.24	.24
Memory	3	.60	.48		.18	.18
Discrimination	4	.30	.24	.18		.09
Cancellation	5	.30	.24	.18	.09	

For instance, let us try the effect of making:

a denote Opposites.
b " Discrimination.
p " Completion.
q " Cancellation.

From the tables of correlations above, we see that r_{ap} will mean the correlation between Opposites and Completion, which is .80. Obtaining in a similar fashion the other three correlations needed, the whole tetrad equation becomes—

$$.80 \times .09 - .30 \times .24 = 0$$

which is obviously correct. And so will be found any other application whatever of the tetrad equation to this table.

The Two Factors. So far, the business is confined to matters of observation; we simply try out the tetrad equation on any table of actually observed correlations and examine whether it fits. The next step, however, is not observational, but purely mathematical; we have to ask how, if at all, this equation between the correlations bears upon the individual measurements of the correlated abilities. The answer is that there has been shown to exist a very remarkable bearing indeed. It is to the effect that, whenever the tetrad equation holds throughout any table of correlations, and *only* when it does so, then every individual measurement of every ability (or of any other variable that enters into the table) can be divided into two independent parts which possess the following momentous properties. The one part has been called the "general factor" and denoted by the letter *g*; it is so named because, although varying freely from individual to individual, it remains the same for any one individual in respect of all the correlated abilities. The second part has been called the "specific factor" and denoted by the letter *s*. It not only varies from individual to individual, but even for any one individual from each ability to another. The proof of this all-important mathematical theorem has gradually evolved through successive stages of completeness, and may now be regarded as complete.

Although, however, both of these factors occur in every ability, they need not be equally influential in all. On the contrary, the very earliest application of this mathematical theorem to psychological correlations showed that there the *g* has a much greater relative influence or "weight" in some of the abilities tested than in others. Means were even found of measuring this relative weight. At one extreme lay the talent for classics, where the ratio of the influence of *g* to that of *s* was rated to be as much as 15 to 1. At the other extreme was the talent for music, where the ratio was only 1 to 4.

Here at once we have before us the essence of the whole doctrine, the seedling from which all else has sprung. But notice must be taken that this general factor *g*, like all measurements anywhere, is primarily not any concrete thing but only a value or magnitude. Further, that which this magnitude measures has not been defined by declaring what it is like, but only by pointing out where it can be found.

It consists in just that constituent—whatever it may be—which is common to all the abilities inter-connected by the tetrad equation. This way of indicating what *g* means is just as definite as when one indicates a card by staking on the back of it without looking at its face.

Such a defining of *g* by site rather than by nature is just what was meant originally when its determination was said to be only "objective." Eventually, we may or may not find reason to conclude that *g* measures something that can appropriately be called "intelligence." Such a conclusion, however, would still never be the definition of *g*, but only a "statement about" it. . . .

Suggested universality of g. The vital problem, in respect of empirical observation, is as to how far and how regularly our tetrad equation actually holds good. In the original work, an extremely wide generalization was adventured. The suggestion was made that

> all branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements seem in every case to be wholly different from that in all the others.

Here, then, lies the justification for attributing so much importance to *g*, despite its purely formal character. The view is put forward that this *g*, far from being confined to some small set of abilities whose inter-correlations have actually been measured and drawn up in some particular table, may enter into all abilities whatsoever.

Such a universal law could only be advanced very tentatively. The express caution was added that

> it must acquire a much vaster corroborative basis before we can accept it even as a general principle and apart from its inevitable corrections and limitations.

This caution was the more imperative, seeing that the law not only had such a tremendous scope, but moreover came into sharp conflict with the most authoritative psychology then prevailing, the latter being at that time wedded to the doctrine of independence.

Utility of the doctrine in practical testing. The preceding doctrine—as we shall see later—admits of usage throughout almost every kind and description of problem within the whole domain of individual difference of ability. For the present, however, we will only allude to two of the most obvious of these applications.

The one consists in the power conferred of measuring any individual in a genuine manner. . . . We can determine the magnitude of his *g* which will tell us nearly everything about some of his abilities and something about nearly all of them. . . . And then we can do the same as regards any of his *s*'s, one for each distinct kind of performance: this supplements and completes the information supplied by his *g*. . . .

The other immediate application of the doctrine is in the construction of mental tests. We are enabled to ascertain just the degree of accuracy with which any given test, or series of them, will measure either a person's *g* or any of his *s*'s. Further, we learn how this degree of accuracy may be raised to its maximum.

We can already see, too, that some crude approach towards measuring *g* can be obtained by the seemingly unscientific course of throwing very miscellaneous tests into a common hotchpot. So doing does not indeed supply an average, or even a representative sample, of the person's abilities; anything of the sort seems to be for ever precluded by the impossibility of fulfilling the necessary postulates. What the pooling does effect is to make the influences of the many specific factors more or less neutralise each other, so that the eventual result will tend to become an approximate measure of *g* alone.

DISCREPANCIES BETWEEN THEORY AND OBSERVATION

Effect of sampling errors. Although we have above suggested that the range of *g* as indicated by the tetrad equation is really universal, this must not be taken to mean that the said equation is under all conditions well satisfied. To expect this would be as absurd as to infer from the law of conservation of energy that the distance a man can walk is always proportional to the amount of food he eats. The manifestations of all such laws, whether mental or physical, are bound to be more or less intermingled with, and modified by, further influences. For such influences, then, all due allowance must be made. In order to verify the law of Two Factors, actual experiment should *not* satisfy the equation exactly, but instead should present exactly the right departure from it. If a marksman wishes to hit the bullseye he does not aim plump at it but more or less to one side, according to the direction and strength of the wind.

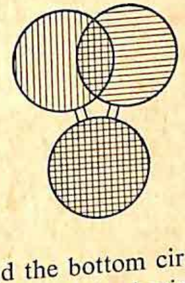
Now, all such inevitable complications of the theorem will be treated in detail later on (ch. x.). But about two of them—to obviate gross misapprehensions—a few words may be said forthwith. One such complication derives from the "errors of sampling." In any actual investigation only a limited number of individuals—ranging generally from 50 to 1,000—can actually be measured; these have to be accepted as a representative sample of the entire class of persons under consideration. But between the correlation found for any such mere sample and that truly holding for the entire class, there must naturally be expected some random discrepancy. The general size of this for any single correlation has long been ascertained; about half of the discrepancies will be greater and about half less than a calculable value called the "probable error" of the correlation. Every table or correlation must, then, be looked upon as consisting of the true values peppered over with, and more or less disguised by, random positive or negative additions: the general size of these, however, will bear a predictable relation to that of the said "probable error."

Such disturbances of the correlations cannot but exercise some effect upon our "tetrad differences," since the latter are constructed out of the former. Consequently, if the true value of the tetrad differences is always zero as shown in the tetrad equation (p. 73), then the actually manifested values ought *not* to be always zero, but instead should present some deviations from this. We need to ascertain how large these deviations should tend to be.

To discover this—or more broadly speaking, to make allowances for the disturbance of the correlations by their errors of sampling—has been the greatest trouble in the whole development of the doctrine. In the earlier investigations, the degree of allowance was left for anybody to estimate as high or as low as he pleased. Such a procedure, of course, is unscientific and misleading. Later on, the artifice was devised of replacing the tetrad equation by another criterion, which would at any rate be approximately correct and which did admit of calculating the allowance to be made for the errors of sampling. This substitute criterion was called that of "inter-columnar correlation." Quite recently, however, this long-standing grave defect in the doctrine has at last been overcome. Means have been discovered for evaluating the effect of the disturbance in the case of the true criterion, the tetrad equation itself. Knowing (as we always do) the probable errors of the four correlations that enter into the tetrad difference (see p. 90) we can now deduce from these the probable error of the tetrad difference as a whole. . . .

Overlap between specific factors. As mentioned, there is another conspicuous limitation to the doctrine we are considering. Obviously, the specific factors for any two performances can only be independent of each other when these performances are quite different. When, on the contrary, two performances are much alike, their respective specific factors will necessarily cease to be mutually independent.

For example, take as one test the cancellation of all the a's in a printed page, and as another test the cancellation of all the e's. These two performances, being so extremely alike, may naturally be expected not only to have the *g* factor in common, but also to present a large overlap in respect of the *s* factor. The case may be symbolized in the following figure, where the verti-



cal shading (top left and the bottom circles) stand for the *s* and the *g* in cancelling a's, whilst the horizontal shading (top right and the bottom circles) stand for the *s* and *g* in cancelling e's. Whenever such cases are introduced, then our tetrad equation must in general *cease* to be valid.

But where, it may be asked, shall the line be drawn between those performances which are and those which are not "quite different." The answer must be that this is not a point to settle intuitively, but to ascertain by experiment. Performances should be regarded as quite different—in the present signification of this phrase—so long as the tetrad equation is satisfied and no longer. To give some notion, however, of how this rule pans out in actual practice, it may at once be said that the most striking instance of such an overlap was afforded by just these two tests that both consist of cancelling letters and only differ as to which letter should be cancelled. Another instance of likeness found to produce marked overlap was that between counting dots one at a time and doing so three at a time. A somewhat different instance was found in school subjects; the correlation between marks for Latin Grammar and those for Latin Translation proved far too large to fit into our tetrad equation.

Evidently, such overlaps are akin to the "common elements" which have been taken by Thorndike, Judd, Sleight, and others to account for any "formal training" that may exist (see ch. iii. p. 36). But our overlaps might also derive, it was found, from various more or less accidental circumstances. Thus, some of the children had been learning Latin for a longer period than the others, and therefore naturally enough tended to excel at both the Grammar and the Translation. By this fact, of course, the correlation between the two abilities must have been increased beyond the amount due to *g* alone.

In general, the early investigations indicated that the cases of appreciable overlap are surprisingly scarce: that is to say, it appeared only to link together abilities that were allied very obviously. But already at this early period an exception was announced. It consisted in the overlap found between the memorizing of syllables and that of numbers. This observation was said to indicate the possibility of a rather extensive group of performances being so nearly related, that they might be gathered together as a more or less unitary ability under the concept of "memorization."

And the point was again emphasized in 1912. Herewith, then, was actually encountered something at least akin to the formal faculties (ch. iii.) or the types (ch. iv.).

After this discovery, such broad factors capable of embracing very extensive ranges of ability were placed in the forefront of enquiry. There has been a steady stream of investigations in our laboratory devoted to ascertaining whether various notable characters of ability do or do not act as specific factors of broad extent. . . .

Overlapping specific factors have since often been spoken of as "group factors." They may be defined as those which occur in more than one but less than all of any given set of abilities. Thus, they indicate no particular characters in any of the abilities themselves, but only some kinship between those which happen to be taken together in a set. Any element whatever in the specific factor of an ability will be turned into a group factor, if this ability is included in the same set with some other ability which also contains this element. The most that can be said is that some elements have a broader range than others, and therefore are more *likely* to play the part of group factors.

9

Three Faces of Intellect

J. P. GUILFORD

The last part of the foregoing selection from Spearman's book points to a research finding that has led to a large number of important investigations in the decades following Spearman's work of the 1920's. The search for group factors, *abilities narrower in their scope than* g *but broader than* s, *became the central focus of attention. A leader in this effort was L. L. Thurstone, who elaborated the concept of* primary mental abilities *and developed techniques to use in exploring them. While Thurstone's reports of his work are too long and involved to be included in this book, J. P. Guilford, who carried forward the kind of research Thurstone began, has produced a synthesis of all the complex factor analytic results that can serve as a definitive theoretical statement about the structure of intellect. The paper is called "Three Faces of Intellect" and is reprinted from* American Psychologist, *1959,* 14, *469–479. It is reprinted with the permission of the author and the American Psychological Association. Introduction and footnotes have been deleted.*

My subject is in the area of human intelligence, in connection with which the names of Terman and Stanford have become known the world over. The Stanford Revision of the Binet intelligence scale has been the standard against which all other instruments for the measurement of intelligence have been compared. The term IQ or intelligence quotient has become a household word in this country. This is illustrated by two brief stories.

A few years ago, one of my neighbors came home from a PTA meeting, remarking: "That Mrs. So-And-So, thinks she knows so much. She kept talking about the 'intelligence *quota*' of the children; 'intelligence *quota*'; imagine. Why, everybody knows that IQ stands for 'intelligence quiz.' "

The other story comes from a little comic strip in a Los Angeles morning newspaper, called "Junior Grade." In the first picture a little boy meets a little girl, both apparently about the first-grade level. The little girl remarks, "I have a high IQ." The little boy, puzzled, said, "You have a what?" The little girl repeated, "I have a high IQ," then went on her way. The little boy, looking thoughtful, said, "And she looks like such a nice little girl, too."

It is my purpose to speak about the analysis of this thing called human intelligence into its components. I do not believe that either Binet or Terman, if they were still with us, would object to the idea of a searching and detailed study of intelligence, aimed toward a better understanding of its nature. Preceding the development of his intelligence scale, Binet had done much research on different kinds of thinking activities and apparently recognized that intelligence has a number of aspects. It is to the lasting credit of both Binet and Terman that they introduced such a great variety of tasks into their intelligence scales.

Two related events of very recent history make it imperative that we learn all we can regarding the nature of intelligence. I am referring to the advent of the artificial satellites and planets and to the crisis in education that has arisen in part as a consequence. The preservation of our way of life and our future security depend upon our most important national resources: our intellectual abilities and, more particularly, our creative abilities. It is time, then, that we learn all we can about those resources.

Our knowledge of the components of human intelligence has come about mostly within the last 25 years. The major sources of this information in this country have been L. L. Thurstone and his associates, the wartime research of psychologists in the United States Air Forces, and more recently the Aptitudes Project at the University of Southern California, now in its tenth year of research on cognitive and thinking abilities. These are mostly novel findings. But to me, the most significant outcome has been the development of a unified theory of human intellect, which organizes the known, unique or primary intellectual abilities into a single system called the "structure of intellect." It is to this system that I shall devote the major part of my remarks, with very brief mentions of some of the implications for the psychology of thinking and problem solving, for vocational testing, and for education.

The discovery of the components of intelligence has been by means of the experimental application of the method of factor analysis. It is not necessary for you to know anything about the theory or method of factor analysis in order to follow the discussion of the components. I should like to say, however, that factor analysis has no connection with or resemblance to psychoanalysis. A positive statement would be more helpful, so I will say that each intellectual component or factor is a unique ability that is needed to do well in a certain class of tasks or tests. As a general principle we find that certain individuals do well in the tests of a certain class, but they may do poorly in the tests of another class. We conclude that a factor has certain properties from the features that the tests of a class have in common. I shall give you very soon a number of examples of tests, each representing a factor.

THE STRUCTURE OF INTELLECT

Although each factor is sufficiently distinct to be detected by factor analysis, in very recent years it has become apparent that the factors themselves can be classified because they resemble one another in certain ways. One basis of classificaton is according to the basic kind of process or operation performed. This kind of classification gives us five major groups of intellectual abilities: factors of cognition, memory, convergent thinking, divergent thinking, and evaluation.

Cognition means discovery or rediscovery or recognition. Memory means retention of what is cognized. Two kinds of productive-thinking operations generate new information from known information and remembered information. In divergent-thinking operations we think in different directions, sometimes searching, sometimes seeking variety. In convergent thinking the information leads to one right answer or to a recognized best or conventional answer. In evaluation we reach decisions as to goodness, correctness, suitability, or adequacy of what we know, what we remember, and what we produce in productive thinking.

A second way of classifying the intellectual factors is according to the kind of material or content involved. The factors known thus far involve three kinds of material or content: the content may be figural, symbolic, or semantic. Figural content is concrete material such as is perceived through the senses. It does not represent anything except itself. Visual material has properties such as size, form, color, location, or texture. Things we hear or feel provide other examples of figural material. Symbolic content is composed of letters, digits, and other conventional signs, usually organized in general systems, such as the alphabet or the number system. Semantic content is in the form of verbal meanings or ideas, for which no examples are necessary.

When a certain operation is applied to a certain kind of content, as many as six general kinds of products may be involved. There is enough evidence available to suggest that, regardless of the combinations of operations and content, the same six kinds of products may be found associated. The six kinds of products are: units, classes, relations, systems, transformations, and implications. So far as we have determined from factor analysis, these are the only fundamental kinds of products that we can know. As such, they may serve as basic classes into which one might fit all kinds of information psychologically.

The three kinds of classifications of the factors of intellect can be represented by means of a single solid model, shown in Figure 1. In this model, which we call the "structure of intellect," each dimension represents one of the modes of variation of the factors. Along one dimension are found the various kinds of operations, along a second one are the various kinds of products, and along the third are various kinds of content. Along the dimension of content a fourth category has been added, its kind of content being designated as "behavioral." This category has been added on a purely theoretical basis to represent the general area sometimes called "social intelligence." More will be said about this section of the model later.

In order to provide a better basis for understanding the model and a better basis for accepting it as a picture of human intellect,

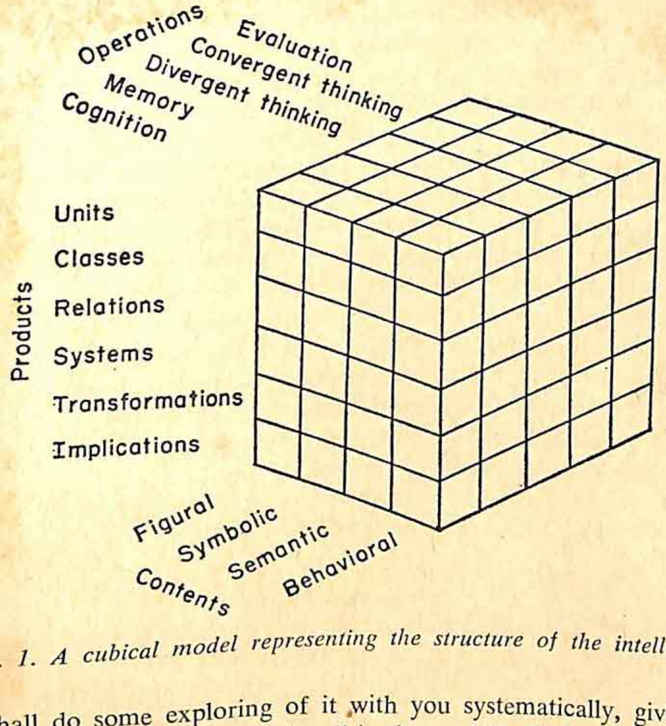


Fig. 1. A cubical model representing the structure of the intellect.

I shall do some exploring of it with you systematically, giving some examples of tests. Each cell in the model calls for a certain kind of ability that can be described in terms of operation, content, and product, for each cell is at the intersection of a unique combination of kinds of operation, content, and product. A test for that ability would have the same three properties. In our exploration of the model, we shall take one vertical layer at a time, beginning with the front face. The first layer provides us with a matrix of 18 cells (if we ignore the behavioral column for which there are as yet no known factors) each of which should contain a cognitive ability.

The Cognitive Abilities

We know at present the unique abilities that fit logically into 15 of the 18 cells for cognitive abilities. Each row presents a triad of similar abilities, having a single kind of product in common. The factors of the first row are concerned with the knowing of units. A good test of the ability to cognize figural units is the Street Gestalt Completion Test. In this test, the recognition of familiar pictured objects in silhouette form is made difficult for testing purposes by blocking out parts of those objects. There is another factor that is known to involve the perception of auditory figures—in the form of melodies, rhythms, and speech sounds—and still another factor involving kinesthetic forms. The presence of three factors in one cell (they are conceivably distinct abilities, although this has not been tested) suggests that more generally, in the figural column, at least, we should expect to find more than one ability. A fourth dimension pertaining to variations in sense modality may thus apply in connection with figural content. The model could be extended in this manner if the facts call for such an extension.

The ability to cognize symbolic units is measured by tests like the following:

Put vowels in the following blanks to make real words:

P__W__R
M__RV__L
C__RT__N

Rearrange the letters to make real words:

R A C I H
T V O E S
K L C C O

The first of these two tests is called Disemvoweled Words, and the second Scrambled Words.

The ability to cognize semantic units is the well-known factor of verbal comprehension, which is best measured by means of a vocabulary test, with items such as:

GRAVITY means ________
CIRCUS means ________
VIRTUE means ________

From the comparison of these two factors it is obvious that recognizing familiar words as letter structures and knowing what words mean depend upon quite different abilities.

For testing the abilities to know classes of units, we may present the following kinds of items, one with symbolic content and one with semantic content:

Which letter group does not belong?
XECM PVAA QXIN VTRO

Which object does not belong?
clam tree oven rose

A figural test is constructed in a completely parallel form, presenting in each item four figures, three of which have a property in common and the fourth lacking that property.

The three abilities to see relationships are also readily measured by a common kind of test, differing only in terms of content. The well-known analogies test is applicable, two items in symbolic and semantic form being:

JIRE : KIRE : : FORA : KORE KORA LIRE GORA GIRE
poetry : prose : : dance : music walk sing talk jump

Such tests usually involve more than the ability to cognize relations, but we are not concerned with this problem at this point.

The three factors for cognizing systems do not at present appear in tests so closely resembling one another as in the case of the examples just given. There is nevertheless an underlying common core of logical similarity. Ordinary space tests, such as Thurstone's Flags, Figures, and Cards or Part V (Spatial Orientation) of the Guilford-Zimmerman Aptitude Survey (GZAS), serve in the figural column. The system involved is an order of arrangement of objects in space. A system that uses symbolic elements is illustrated by the Letter Triangle Test, a sample item of which is:

—
d —
b e —
a c f ?

What letter belongs at the place of the question mark?

The ability to understand a semantic system has been known for some time as the factor called general reasoning. One of its most faithful indicators is a test composed of arithmetic-reasoning items. That the phase of understanding only is important for measuring this ability is shown by the fact that such a test works even if the examinee is not asked to give a complete solution; he need only show that he structures the problem properly. For example, an item from the test Necessary Arithmetical Operations simply asks what operations are needed to solve the problem:

A city lot 48 feet wide and 149 feet deep costs $79,432. What is the cost per square foot?

A. add and multiply
B. multiply and divide
C. subtract and divide
D. add and subtract
E. divide and add

Placing the factor of general reasoning in this cell of the structure of intellect gives us some new conceptions of its nature. It should be a broad ability to grasp all kinds of systems that are conceived in terms of verbal concepts, not restricted to the understanding of problems of an arithmetical type.

Transformations are changes of various kinds, including modifications in arrangement, organization, or meaning. In the figural column for the transformations row, we find the factor known as visualization. Common measuring instruments for this factor are the surface-development tests, and an example of a different kind is Part VI (Spatial Visualization) of the GZAS. A test of the ability to make transformations of meaning, for the factor in the semantic column, is called Similarities. The examinee is asked to state several ways in which two objects, such as an apple and an orange, are alike. Only by shifting the meanings of both is the examinee able to give many responses to such an item.

In the set of abilities having to do with the cognition of implications, we find that the individual goes beyond the information given, but not to the extent of what might be called drawing conclusions. We may say that he extrapolates. From the given information he expects or foresees certain consequences, for example. The two factors found in this row of the cognition matrix were first called "foresight" factors. Foresight in connection with figural material can be tested by means of paper-and-pencil mazes. Foresight in connection with ideas, those pertaining to events, for example, is indicated by a test such as Pertinent Questions:

> In planning to open a new hamburger stand in a certain community, what four questions should be considered in deciding upon its location?

The more questions the examinee asks in response to a list of such problems, the more he evidently foresees contingencies.

The Memory Abilities

The area of memory abilities has been explored less than some of the other areas of operation, and only seven of the potential cells of the memory matrix have known factors in them. These cells are restricted to three rows: for units, relations, and systems. The first cell in the memory matrix is now occupied by two factors, parallel to two in the corresponding cognition matrix: visual memory and auditory memory. Memory for series of letters or numbers, as in memory span tests, conforms to the conception of memory for symbolic units. Memory for the ideas in a paragraph conforms to the conception of memory for semantic units.

The formation of associations between units, such as visual forms, syllables, and meaningful words, as in the method of paired associates, would seem to represent three abilities to remember relationships involving three kinds of content. We know of two such abilities, for the symbolic and semantic columns. The memory for known systems is represented by two abilities very recently discovered (Christal, 1958). Remembering the arrangement of objects in space is the nature of an ability in the figural column, and remembering a sequence of events is the nature of a corresponding ability in the semantic column. The differentiation between these two abilities implies that a person may be able to say where he saw an object on a page, but he might not be able to say on which of several pages he saw it after leafing through several pages that included the right one. Considering the blank rows in the memory matrix, we should expect to find abilities also to remember classes, transformations, and implications, as well as units, relations, and systems.

The Divergent-Thinking Abilities

The unique feature of divergent production is that a *variety* of responses is produced. The product is not completely determined by the given information. This is not to say that divergent thinking does not come into play in the total process of reaching a unique conclusion, for it comes into play wherever there is trial-and-error thinking.

The well-known ability of word fluency is tested by asking the examinee to list words satisfying a specified letter requirement, such as words beginning with the letter "s" or words ending in "-tion." This ability is now regarded as a facility in divergent production of symbolic units. The parallel semantic ability has been known as ideational fluency. A typical test item calls for listing objects that are round and edible. Winston Churchill must have possessed this ability to a high degree. Clement Attlee is reported to have said about him recently that, no matter what problem came up, Churchill always seemed to have about ten ideas. The trouble was, Attlee continued, he did not know which was the good one. The last comment implies some weakness in one or more of the evaluative abilities.

The divergent production of class ideas is believed to be the unique feature of a factor called "spontaneous flexibility." A typical test instructs the examinee to list all the uses he can think of for a common brick, and he is given eight minutes. If his responses are: build a house, build a barn, build a garage, build a school, build a church, build a chimney, build a walk, and build a barbeque, he would earn a fairly high score for ideational fluency but a very low score for spontaneous flexibility, because all these uses fall into the same class. If another person said: make a door stop, make a paper weight, throw it at a dog, make a bookcase, drown a cat, drive a nail, make a red powder, and use for baseball bases, he would also receive a high score for flexibility. He has gone frequently from one class to another.

A current study of unknown but predicted divergent-production abilities includes testing whether there are also figural and symbolic abilities to produce multiple classes. An experimental figural test presents a number of figures that can be classified in groups of three in various ways, each figure being usable in more than one class. An experimental symbolic test presents a few numbers that are also to be classified in multiple ways.

A unique ability involving relations is called "associational fluency." It calls for the production of a variety of things related in a specified way to a given thing. For example, the examinee is asked to list words meaning about the same as "good" or to list words meaning about the opposite of "hard." In these instances the response produced is to complete a relationship, and semantic content is involved. Some of our present experimental tests call for the production of varieties of relations, as such, and involve figural and symbolic content also. For example, given four small digits, in how many ways can they be related in order to produce a sum of eight?

One factor pertaining to the production of systems is known as expressional fluency. The rapid formation of phrases or sentences is the essence of certain tests of this factor. For example, given the initial letters:

W_____ c_____ e_____ n_____

with different sentences to be produced, the examinee might write "We can eat nuts" or "Whence came Eve Newton?" In interpreting the factor, we regard the sentence as a symbolic system. By analogy, a figural system would be some kind of organization of lines and other elements, and a semantic system would be in the form of a verbally stated problem or perhaps something as complex as a theory.

In the row of the divergent-production matrix devoted to transformations, we find some very interesting factors. The one called "adaptive flexibility" is now recognized as belonging in the figural column. A faithful test of it has been Match Problems. This is based upon the common game that uses squares, the sides of which are formed by match sticks. The examinee is told to take away a given number of matches to leave a stated number of squares with nothing left over. Nothing is said about the sizes of the squares to be left. If the examinee imposes upon himself the restriction that the squares that he leaves must be of the same size, he will fail in his attempts to do items like that in Figure 2. Other odd kinds of solutions are introduced in other items, such as overlapping squares and squares within squares, and so on. In another variation of Match Problems the examinee is told to produce two or more solutions for each problem.

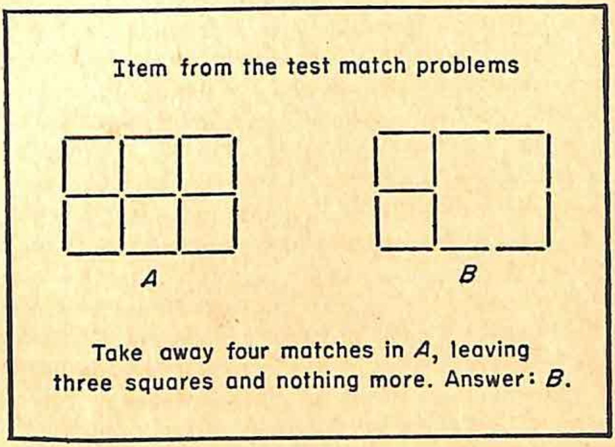

Item from the test match problems
A
B
Take away four matches in A, leaving three squares and nothing more. Answer: B.

Fig. 2. A sample item from the test Match Problems. The problem in this item is to take away four matches and have three squares. The solution is given.

A factor that has been called "originality" is now recognized as adaptive flexibility with semantic material, where there must be a shifting of meanings. The examinee must produce the shifts or changes in meaning and so come up with novel, unusual, clever, or farfetched ideas. The Plot Titles Test presents a short story, the examinee being told to list as many appropriate titles as he can to head the story. One story is about a missionary who has been captured by cannibals in Africa. He is in the pot and about to be boiled when a princess of the tribe obtains a promise for his release if he will become her mate. He refuses and is boiled to death.

In scoring the test, we separate the responses into two categories, clever and nonclever. Examples of nonclever responses are: African Death, Defeat of a Princess, Eaten by Savages, The Princess, The African Missionary, In Darkest Africa, and Boiled by Savages. These titles are appropriate but commonplace. The number of such responses serves as a score for ideational fluency. Examples of clever responses are: Pot's Plot, Potluck Dinner, Stewed Parson, Goil or Boil, A Mate Worse than Death, He Left a Dish for a Pot, Chaste in Haste, and A Hot Price for Freedom. The number of clever responses given by an examinee is his score for originality, or the divergent production of semantic transformations.

Another test of originality presents a very novel task so that any acceptable response is unusual for the individual. In the Symbol Production Test the examinee is to produce a simple symbol to stand for a noun or a verb in each short sentence, in other words to invent something like pictographic symbols. Still another test of originality asks for writing the "punch lines" for cartoons, a task that almost automatically challenges the examinee to be clever. Thus, quite a variety of tests offer approaches to the measurement of originality, including one or two others that I have not mentioned.

Abilities to produce a variety of implications are assessed by tests calling for elaboration of given information. A figural test of this type provides the examinee with a line or two, to which he is to add other lines to produce an object. The more lines he adds, the greater his score. A semantic test gives the examinee the outlines of a plan to which he is to respond by stating all the details he can think of to make the plan work. A new test we are trying out in the symbolic area presents two simple equations such as $B - C = D$ and $z = A + D$. The examinee is to make as many other equations as he can from this information.

The Convergent-Production Abilities

Of the 18 convergent-production abilities expected in the three content columns, 12 are now recognized. In the first row, pertaining to units, we have an ability to name figural properties (forms or colors) and an ability to name abstractions (classes, relations, and so on). It may be that the ability in common to the speed of naming forms and the speed of naming colors is not appropriately placed in the convergent-thinking matrix. One might expect that the thing to be produced in a test of the convergent production of figural units would be in the form of figures rather than words. A better test of such an ability might somehow specify the need for one particular object, the examinee to furnish the object.

A test for the convergent production of classes (Word Grouping) presents a list of 12 words that are to be classified in four, and only four, meaningful groups, no word to appear in more than one group. A parallel test (Figure Concepts Test) presents 20 pictured real objects that are to be grouped in meaningful classes of two or more each.

Convergent production having to do with relationships is represented by three known factors, all involving the "eduction of correlates," as Spearman called it. The given information includes one unit and a stated relation, the examinee to supply the other unit. Analogies tests that call for completion rather than a choice between alternative answers emphasize this kind of ability. With symbolic content such an item might read:

pots stop bard drab rats .?

A semantic item that measures eduction of correlates is:

The absence of sound is ________.

Incidentally, the latter item is from a vocabulary-completion test, and its relation to the factor of ability to produce correlates indicates how, by change of form, a vocabulary test may indicate an ability other than that for which vocabulary tests are usually intended, namely, the factor of verbal comprehension.

Only one factor for convergent production of systems is known, and it is in the semantic column. It is measured by a class of tests that may be called ordering tests. The examinee may be presented with a number of events that ordinarily have a best or most logical order, the events being presented in scrambled order. The presentation may be pictorial, as in the Picture Arrangement Test, or verbal. The pictures may be taken from a cartoon strip. The verbally presented events may be in the form of the various steps needed to plant a new lawn. There are undoubtedly other kinds of systems than temporal order that could be utilized for testing abilities in this row of the convergent-production matrix.

In the way of producing transformations of a unique variety, we have three recognized factors, known as redefinition abilities. In each case, redefinition involves the changing of functions or uses of parts of one unit and giving them new functions or uses in some new unit. For testing the ability of figural redefinition, a task based upon the Gottschaldt figures is suitable. Figure 3 shows the kind of item for such a test. In recognizing the simpler figure within the structure of a more complex figure, certain lines must take on new roles.

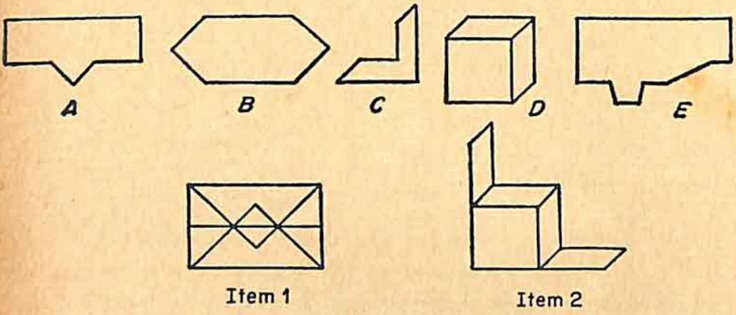


Fig. 3. Sample items from a test Hidden Figure, based upon the Gottschaldt figures. Which of the simple figures is concealed within each of the two more complex figures?

In terms of symbolic material, the following sample items will illustrate how groups of letters in given words must be readapted to use in other words. In the test Camouflaged Words, each sentence contains the name of a sport or game:

I did not know that he was ailing.
To beat the Hun, tin goes a long way.

For the factor of semantic redefinition, the Gestalt Transformation Test may be used. A sample item reads:

From which object could you most likely make a needle?
A. a cabbage
B. a splice

C. a steak
D. a paper box
E. a fish

The convergent production of implications means the drawing of fully determined conclusions from given information. The well-known factor of numerical facility belongs in the symbolic column. For the parallel ability in the figural column, we have a test known as Form Reasoning, in which rigorously defined operations with figures are used. For the parallel ability in the semantic column, the factor sometimes called "deduction" probably qualifies. Items of the following type are sometimes used.

Charles is younger than Robert
Charles is older than Frank
Who is older: Robert or Frank?

Evaluative Abilities

The evaluative area has had the least investigation of all the operational categories. In fact, only one systematic analytical study has been devoted to this area. Only eight evaluative abilities are recognized as fitting into the evaluation matrix. But at least five rows have one or more factors each, and also three of the usual columns or content categories. In each case, evaluation involves reaching decisions as to the accuracy, goodness, suitability, or workability of information. In each row, for the particular kind of product of that row, some kind of criterion or standard of judgment is involved.

In the first row, for the evaluation of units, the important decision to be made pertains to the identity of a unit. Is this unit identical with that one? In the figural column we find the factor long known as "perceptual speed." Tests of this factor invariably call for decisions of identity, for example, Part IV (Perceptual Speed) of the GZAS or Thurstone's Identical Forms. I think it has been generally wrongly thought that the ability involved is that of cognition of visual forms. But we have seen that another factor is a more suitable candidate for this definition and for being in the very first cell of the cognitive matrix. It is parallel to this evaluative ability but does not require the judgment of identity as one of its properties.

In the symbolic column is an ability to judge identity of symbolic units, in the form of series of letters or numbers or of names of individuals.

Are members of the following pairs identical or not:
825170493________825176493
dkeltvmpa________dkeltvmpa
C. S. Meyerson________C. E. Meyerson

Such items are common in tests of clerical aptitude.

There should be a parallel ability to decide whether two ideas are identical or different. Is the idea expressed in this sentence the same as the idea expressed in that one? Do these two proverbs express essentially the same idea? Such tests exist and will be used to test the hypothesis that such an ability can be demonstrated.

No evaluative abilities pertaining to classes have as yet been recognized. The abilities having to do with evaluation where relations are concerned must meet the criterion of logical consistency. Syllogistic-type tests involving letter symbols indicate a different ability than the same type of test involving verbal statements. In the figural column we might expect that tests incorporating geometric reasoning or proof would indicate a parallel ability to sense the soundness of conclusions regarding figural relationships.

The evaluation of systems seems to be concerned with the internal consistency of those systems, so far as we can tell from the knowledge of one such factor. The factor has been called "experiential evaluation," and its representative test presents items like that in Figure 4 asking "What is wrong with this picture?" The things wrong are often internal inconsistencies.

A semantic ability for evaluating transformations is thought to be that known for some time as "judgment." In typical judgment tests, the examinee is asked to tell which of five solutions to a practical problem is most adequate or wise. The solutions frequently involve improvisations, in other words, adaptations of

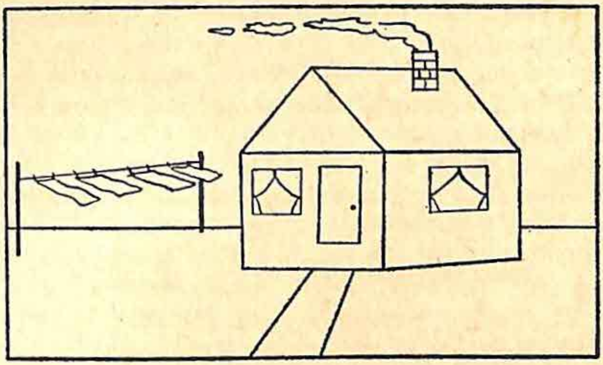


Fig. 4.

familiar objects to unusual uses. In this way the items present redefinitions to be evaluated.

A factor known first as "sensitivity to problems" has become recognized as an evaluative ability having to do with implications. One test of the factor, the Apparatus Test, asks for two needed improvements with respect to each of several common devices, such as the telephone or the toaster. The Social Institutions Test, a measure of the same factor, asks what things are wrong with each of several institutions, such as tipping or national elections. We may say that defects or deficiencies are implications of an evaluative kind. Another interpretation would be that seeing defects and deficiencies are evaluations of implications to the effect that the various aspects of something are all right.

SOME IMPLICATIONS OF THE STRUCTURE OF INTELLECT

For Psychological Theory

Although factor analysis as generally employed is best designed to investigate ways in which individuals differ from one another, in other words, to discover traits, the results also tell us much about how individuals are alike. Consequently, information regarding the factors and their interrelationships gives us understanding of functioning individuals. The five kinds of intellectual abilities in terms of operations may be said to represent five ways of functioning. The kinds of intellectual abilities distinguished according to varieties of test content and the kinds of abilities distinguished according to varieties of products suggest a classification of basic forms of information or knowledge. The kind of organism suggested by this way of looking at intellect is that of an agency for dealing with information of various kinds in various ways. The concepts provided by the distinctions among the intellectual abilities and by their classifications may be very useful in our future investigations of learning, memory, problem solving, invention, and decision making, by whatever method we choose to approach those problems.

For Vocational Testing

With about 50 intellectual factors already known, we may say that there are at least 50 ways of being intelligent. It has been facetiously suggested that there seem to be a great many more ways of being stupid, unfortunately. The structure of intellect is a theoretical model that predicts as many as 120 distinct abilities, if every cell of the model contains a factor. Already we know that two cells contain two or more factors each, and there probably are actually other cells of this type. Since the model was first conceived, 12 factors predicted by it have found places in it. There is consequently hope of filling many of the other vacancies, and we may eventually end up with more than 120 abilities.

The major implication of the assessment of intelligence is that to know an individual's intellectual resources thoroughly we shall need a surprisingly large number of scores. It is expected that many of the factors are intercorrelated, so there is some possibility that by appropriate sampling we shall be able to cover the important abilities with a more limited number of tests. At any rate, a multiple-score approach to the assessment of intelligence is definitely indicated in connection with future vocational operations.

Considering the kinds of abilities classified as to content, we may speak roughly of four kinds of intelligence. The abilities involving the use of figural information may be regarded as "concrete" intelligence. The people who depend most upon these abilities deal with concrete things and their properties. Among these people are mechanics, operators of machines, engineers (in some aspects of their work), artists, and musicians.

In the abilities pertaining to symbolic and semantic content, we have two kinds of "abstract" intelligence. Symbolic abilities should be important in learning to recognize words, to spell, and to operate with numbers. Language and mathematics should depend very much upon them, except that in mathematics some aspects, such as geometry, have strong figural involvement. Semantic intelligence is important for understanding things in terms of verbal concepts and hence is important in all courses where the learning of facts and ideas is essential.

In the hypothesized behavioral column of the structure of intellect, which may be roughly described as "social" intelligence, we have some of the most interesting possibilities. Understanding the behavior of others and of ourselves is largely nonverbal in character. The theory suggests as many as 30 abilities in this area, some having to do with understanding, some with productive thinking about behavior, and some with the evaluation of behavior. The theory also suggests that information regarding behavior is also in the form of the six kinds of products that apply elsewhere in the structure of intellect, including units, relations, systems, and so on. The abilities in the area of social intelligence, whatever they prove to be, will possess considerable importance in connection with all those individuals who deal most with other people: teachers, law officials, social workers, therapists, politicians, statesmen, and leaders of other kinds.

For Education

The implications for education are numerous, and I have time just to mention a very few. The most fundamental implication is that we might well undergo transformations with respect to our conception of the learner and of the process of learning. Under the prevailing conception, the learner is a kind of stimulus-response device, much on the order of a vending machine. You put in a coin, and something comes out. The machine learns what reaction to put out when a certain coin is put in. If, instead, we think of the learner as an agent for dealing with information, where information is defined very broadly, we have something more analogous to an electronic computer. We feed a computer information; it stores that information; it uses that information for generating new information, either by way of divergent or convergent thinking; and it evaluates its own results. Advantages that a human learner has over a computer include the step of seeking and discovering new information from sources outside itself and the step of programing itself. Perhaps even these steps will be added to computers, if this has not already been done in some cases.

At any rate, this conception of the learner leads us to the idea that learning is discovery of information, not merely the formation of associations, particularly associations in the form of stimulus-response connections. I am aware of the fact that my proposal is rank heresy. But if we are to make significant progress in our understanding of human learning and particularly our understanding of the so-called higher mental processes of thinking, problem solving, and creative thinking, some drastic modifications are due in our theory.

The idea that education is a matter of training the mind or of training the intellect has been rather unpopular, wherever the prevailing psychological doctrines have been followed. In theory, at least, the emphasis has been upon the learning of rather specific habits or skills. If we take our cue from factor theory, however, we recognize that most learning probably has both specific and general aspects or components. The general aspects may be along the lines of the factors of intellect. This is not to say that the individual's status in each factor is entirely determined by learning. We do not know to what extent each factor is determined by heredity and to what extent by learning. The best position for educators to take is that possibly every intellectual factor can be developed in individuals at least to some extent by learning.

If education has the general objective of developing the intellects of students, it can be suggested that each intellectual factor provides a particular goal at which to aim. Defined by a certain combination of content, operation, and product, each goal ability then calls for certain kinds of practice in order to achieve improvement in it. This implies choice of curriculum and the choice or invention of teaching methods that will most likely accomplish the desired results.

Considering the very great variety of abilities revealed by the factorial exploration of intellect, we are in a better position to ask whether any general intellectual skills are now being neglected in education and whether appropriate balances are being observed. It is often observed these days that we have fallen down in the way of producing resourceful, creative graduates. How true this is, in comparison with other times, I do not know. Perhaps the deficit is noticed because the demands for inventiveness are so much greater at this time. At any rate, realization that the more conspicuously creative abilities appear to be concentrated in the divergent-thinking category, and also to some extent in the transformation category, we now ask whether we have been giving these skills appropriate exercise. It is probable that we need a better balance of training in the divergent-thinking area as compared with training in convergent thinking and in critical thinking or evaluation.

The structure of intellect as I have presented it to you may or may not stand the test of time. Even if the general form persists, there are likely to be some modifications. Possibly some different kind of model will be invented. Be that as it may, the fact of a multiplicity of intellectual abilities seems well established.

There are many individuals who long for the good old days of simplicity, when we got along with one unanalyzed intelligence. Simplicity certainly has its appeal. But human nature is exceedingly complex, and we may as well face that fact. The rapidly moving events of the world in which we live have forced upon us the need for knowing human intelligence thoroughly. Humanity's peaceful pursuit of happiness depends upon our control of nature and of our own behavior; and this, in turn, depends upon understanding ourselves, including our intellectual resources.

REFERENCES

CHRISTAL, R. E. Factor analytic study of visual memory. *Psychol. Monogr.*, 1958, 72, No. 13 (Whole No. 466).

GUILFORD, J. P. The structure of intellect. *Psychol. Bull.*, 1956, 53, 267–293.

GUILFORD, J. P. *Personality*. New York: McGraw-Hill, 1959.

10

The Study of Giftedness: A Multidimensional Approach

J. W. GETZELS AND P. W. JACKSON

During the 1950's and 1960's increasing attention has been focused on the question of whether creativity *is a characteristic different from and independent of intelligence as usually tested. One widely discussed investigation was that of J. W. Getzels and P. W. Jackson. They reported their results in full in book form, but a brief account of what they did and what they found was published in the paper from which portions are reprinted here. The reference is "The Study of Giftedness: A Multidimensional Approach," in* The Gifted Student, *OE-35016, Cooperative Research Monograph No. 2. The sections below are from pages 6–18. They are reprinted with the authors' permission.*

THE HIGHLY CREATIVE AND THE HIGHLY INTELLIGENT ADOLESCENT: *An Attempt to Differentiate Two Types of Giftedness*

Although few would argue that the terms "creative" and "intelligent" refer to independent classes of phenomena, still fewer would seriously propose that the two terms are synonymous. Despite this recognition of relative independence at the theoretical level, when turning to practical problems involving the assessment of the two qualities, one finds that the well-established intelligence test is most frequently used as *the* indicator of creativity. Individuals are grouped by IQ, and generally there is disappointment if the high-IQ person is not also creative. Sometimes he is not. There is also bewilderment when the creative individual has a relatively low IQ, as is often the case. Theoretical issues in the study of higher cognitive functioning and current practical problems in the search for talent make the importance of differentiating the two concepts in empirical terms self-evident.

Problem

The central purpose of this research is to discover significant variables differentiating the creative from the intelligent person. Specifically, the present study examines the achievement motives, fantasy productions, school performance, and teacher preferences of two types of adolescents: those exceptionally high in creativity but not in IQ, and those exceptionally high in IQ, but not in creativity.

Subjects, Instruments, and Procedures

The subjects were drawn from 449 adolescents comprising the total population of a midwestern private secondary school. The experimental groups were composed on the basis of performance on the following tests:

Standard IQ tests.—A Stanford-Binet, a WISC, or a Henmon-Nelson score was available for each adolescent. The scores obtained from the WISC and the Henmon-Nelson were converted by regression equation to comparable Stanford-Binet IQ's.

Five creativity tests taken or adapted from Guilford and Cattell or constructed especially for the study:

1. WORD ASSOCIATION. The subject was asked to give as many definitions as possible to fairly common stmulus-words; e.g., "bolt," "bark," "sack." His score depended upon the absolute number of definitions and the number of different categories into which these definitions could be put.
2. USES FOR THINGS. The subject was required to give as many uses as he could for objects that customarily have a stereotyped function attached to them; e.g., "brick," "paperclip." His score depended upon both the number and originality of the uses he mentioned.

3. HIDDEN SHAPES. The subject was required to find a given geometric form that was hidden in more complex geometric forms or patterns.
4. FABLES. The subject was presented with four fables in which the last lines were missing. He was required to compose three different endings for *each* fable: a "moralistic," a "humorous," and a "sad" ending. His score depended upon the number, appropriateness, and originality of the endings.
5. MAKE-UP PROBLEMS. The subject was presented with four complex paragraphs, each of which contained a number of numerical statements; e.g., "the costs involved in building a house." He was asked to make all the mathematical problems he could that might be solved with the information given. His score depended upon the number, appropriateness, and originality of the problems.

On the basis of the IQ measure and the mean of the five creativity measures, two experimental groups were formed:

1. *The high creativity group* (26 students: 15 boys, 11 girls) —students in the top 20 percent in creativity, compared to students of the same age and sex, but not in the top 20 percent in IQ.
2. *The high intelligence group* (28 students: 17 boys, 11 girls) —students in the top 20 percent in IQ, compared with students of the same age and sex, but not in the top 20 percent in creativity.

The two groups were then compared to each other and to the population from which they were drawn on the following variables: School performance as measured by standardized verbal and numerical achievement tests appropriate to each grade level; teacher preferences on having them in class; N-Achievement as measured by conventional scoring of responses to six of McClelland's stimulus-pictures. In addition to the conventional scoring, the N-Achievement protocols were also sorted "blind" for the creative and a matched noncreative group and finally analyzed systematically for the creative and intelligent groups by categories suggested in the blind sorting.

Results

The results (see Tables 1 and 2) may be summarized as follows: (1) Despite the striking differences in mean IQ, the creative and the intelligent groups were *equally* superior to the total population in school performance as measured by standardized achievement tests. (2) Although it might be expected that creativity would contribute to status in a school setting, especially since the achievement of the highly creative subjects was at

TABLE 1

Means and Standard Deviations of Highly Creative and Highly Intelligent Groups on Experimental Variables

Variables	Total population[1] ($N = 449$)	High-IQ ($N = 28$)	High creative[2] ($N = 24$)
1	2	3	4
IQ			
Mean	132.00	150.00**	127.00
Standard deviation	15.07	6.64	10.58
School achievement			
Mean	49.91	55.00**	56.27**
Standard deviation	7.36	5.95	7.90
Teacher preference ratings			
Mean	10.23	11.20*	10.54
Standard deviation	3.64	1.56	1.95
Need for achievement (T.-scores):			
Mean	49.81	49.00	50.04
Standard deviation	9.49	7.97	8.39

[1] For purposes of comparison the scores of each experimental group were extracted from the total population before *t*-tests were computed.

[2] Two subjects were omitted from the sample of 26 high creatives because because they failed to furnish data on the McClelland *N*-Achievement instrument.

* Significant at 0.01 level.

** Significant at 0.001 level.

least as good as that of the highly intelligent subjects, it was actually the *intelligent* group rather than the *creative* group that was preferred by teachers when compared with the average student. (3) The two groups did *not* differ from the population in N-Achievement as scored conventionally. (4) Although failing to show differences when scored for N-Achievement, the fantasy productions of the creative group were clearly dissimilar from those of the population as a whole, so much so that the "blind" sorting of 47 creative and noncreative protocols resulted in only 7 misplacements. (5) A more systematic differentiation of the protocols using content-analysis procedures showed that the two experimental groups differed significantly in their fantasy responses. The creative group made greater use of stimulus-free themes, unexpected endings, humor, incongruities, and playfulness.

Here, for example, are typical stories by highly intelligent and by highly creative subjects to two of the McClelland stimulus-pictures.

TABLE 2

Categories of Fantasy Production of Highly Creative and Highly Intelligent Groups

	Group				
	High creativity[1] ($N = 24$)		High IQ ($N = 28$)		
Content-analysis categories	Frequency[2]	Percent	Frequency[2]	Percent	χ^2
1	2	3	4	5	6
Stimulus-free theme	18	75	11	39	5.31[3]*
Unexpected ending	22	92	17	61	5.05*
Presence of humor	17	71	7	25	9.16**
Presence of incongruity	17	71	10	36	5.06*
Presence of violence	18	75	13	46	3.27
Playful attitude toward theme	21	89	9	32	14.04**

[1] Two subjects were omitted from the sample of 26 High Creatives because they failed to furnish data on the McClelland N-Achievement instrument.

[2] Numbers in the frequency column represent the subjects whose fantasy productions fit the corresponding categories.

[3] Yates correction was applied in the computation of chi-squares.

* Significant at the 0.05 level.

** Significant at the 0.01 level.

One picture-stimulus was perceived most often as a man in an airplane reclining seat returning from a business trip or conference. A high-IQ subject gave the following story: "Mr. Smith is on his way home from a successful business trip. He is very happy and he is thinking about his wonderful family and how glad he will be to see them again. He can picture it, about an hour from now, his plane landing at the airport and Mrs. Smith and their three children all there welcoming him home again." A high-creative subject wrote this story: "This man is flying back from Reno where he has just won a divorce from his wife. He couldn't stand to live with her anymore, he told the judge, because she wore so much cold cream on her face at night that her head would skid across the pillow and hit him in the head. He is now contemplating a new skid-proof face cream."

This story was in response to a stimulus-picture usually perceived as a man working late (or very early) in an office:

THE HIGH IQ STUDENT: "There's ambitious Bob, down at the office at 6:30 in the morning. Every morning it's the same. He's trying to show his boss how energetic he is. Now, thinks Bob, maybe the boss will give me a raise for all my extra work. The trouble is that Bob has been doing this for the last three years, and the boss still hasn't given him a raise. He'll come in at 9:00, not even noticing that Bob had been there so long, and poor Bob won't get his raise."

THE HIGH-CREATIVE STUDENT: "This man has just broken into this office of a new cereal company. He is a private-eye employed by a competitor firm to find out the formula that makes the cereal bend, sag, and sway. After a thorough search of the office he comes upon what he thinks is the correct formula. He is now copying it. It turns out that it is the wrong formula and the competitor's factory blows up. Poetic justice!"

These stories were written in group sessions with often more than a hundred adolescents in the same room and a maximum writing time four minutes per story. "Skidproof face cream!" "Cereal that will bend, sag, and sway!" It seems that the ability to restructure stereotyped objects with ease and rapidity—almost "naturally"—is the characteristic mark of the high creative as against the high IQ subjects in this investigation.

Discussion and Summary

The essence of creativity appears to lie in the ability to produce new forms, to conjoin elements that are customarily thought of as independent or dissimilar; not merely the propensity for seeing the bizarre but rather the aptitude for achieving new meanings having social value. Guilford has given the name "divergent" to intellectual activities expressing this ability, in contrast to activities which he calls "convergent." It is chiefly the latter type of performance that is tapped by conventional measures of intelligence. Whether or not one adopts the terminology of Guilford, there is little doubt that the ability required for what has been called "divergent" intellectual activity is not well sampled by the usual intelligence-test items, and accordingly this ability contributes but little to the IQ score. In these terms, the high IQ subjects of the experimental groups are superior in "convergent" intellectual performance and the high-creativity subjects in "divergent" intellectual performance.

The evidence of this report centers attention on the difference in the fantasy quality of the two groups. It is tentatively concluded that a *sine qua non* of creative functioning is a rich and available fantasy life.

The difference between the person with a convergent and the person with a divergent orientation toward problem solving appears to be that the former focuses on the stimulus, since he seeks objective precision of the sort others will recognize as "right" and therefore successful by conventional standards. In effect, he may be said to intellectualize the ambiguities presented by a problem in an effort to find the right answer. The latter detaches himself from the stimulus, since he prefers subjective playfulness that he himself will find "delightful." In effect, he may be said to personalize the ambiguities presented by a problem in an effort to attain an enjoyment which may or may not be shared with others. If the former "latches on" *to* a stimulus, the latter "takes off" *from* a stimulus.

The search for variables differentiating the "creative" from the "intelligent" person is the focus of much theoretical and practical concern. This study examined the achievement motives, fantasy productions, school performance, and teacher preferences of two types of adolescents, those exceptionally high in creativity but not in IQ and those exceptionally high in IQ but not in creativity. The results indicated that: (1) Despite striking differences in IQ, the two groups were equally superior in school achievement; (2) The intelligent subjects rather than the creative ones were preferred by teachers; (3) There were no differences in N-Achievement; (4) There were significant differences in fantasy productions, the creative group using more stimulus-free, humorous, and playful themes.

OCCUPATIONAL CHOICE AND COGNITIVE FUNCTIONING: *A Study of the Career Aspirations of Highly Intelligent and of Highly Creative Adolescents*

The relationship between an individual's cognitive style and his personal orientation to the world is of central importance to the understanding of both cognitive functioning and social behavior. This study investigates the relationship between cognition, as defined by performance on intellective tasks, and personal orientation, as defined by long-term occupational choice and career aspiration.

Problem

The purpose of the investigation is two-fold: At the substantive level, to determine the occupational preferences of the two groups of adolescents described in the previous section; at the theoretical level, to ascertain the relationship between types of cognition and types of career aspiration. Two guiding empirical questions were posed: (1) What is the nature of the long-term occupational goals of the two groups? (2) To what extent do perceived "teacher standards" and the criterion "success in adult life" enter as motives in the aspirations of the two groups?

Subjects, Instruments, and Procedures

The experimental groups were the 26 highly creative and the 28 highly intelligent adolescents previously described. Four instruments were used to obtain the data:

A DIRECT SENTENCE COMPLETION TEST. Among the 60 items comprising this test were such career-preference items as: "When I grow up I want to become a . . . .," "If the choice of a permanent occupation were left up to me, I would choose . . . ," "The work that has the greatest appeal to me is . . . ," "I always wanted to be a . . ." The student was required to respond directly to these items by completing each stub so as to reflect his own preferences. The responses of the two groups were analyzed for the number and kind of occupations mentioned.

AN INDIRECT SENTENCE COMPLETION TEST. This test was parallel to the direct test, except that the sentence stubs were written in the third person and used boys' names for the male subjects and girls' names for the female subjects: "When Stan [Jean] grows up, he [she], wants to become a . . . ," etc. The subject was required to respond "projectively" to the items by completing the stubs as a verbal speed test. The responses of the two groups were analyzed for the number and kind of occupations mentioned.

A PARENT QUESTIONNAIRE. This instrument required the parents of the students to respond to two questions regarding their children's occupational goals as follows: "Has the pupil expressed interest in a particular career?" "If yes, specify. . . ." The responses from the parents of the two groups were analyzed for the number and kind of occupations mentioned.

AN OUTSTANDING TRAITS TEST. This was the same test described in detail in the first section. It contained descriptions of 13 children, each description exemplifying some desirable personal quality or trait. The subjects were required to rank the 13 hypothetical children in 3 ways: (1) On the degree to which they would like to be like them; (2) On the degree to which they believed teachers would like them; (3)

on the degree to which they believed people with these qualities would succeed in adult life. The rankings given by the two groups within each category and the relationships among the rankings for the different categories were analyzed.

Summary of Results

1. Differences between the two groups appeared in both the *quantity* and *quality* of occupational goals. The *quantity* of occupational possibilities mentioned is significantly greater for the high creatives than for the high IQ's, the latter group having mentioned an average of 1.82 different occupations on the Direct Sentence Completion Test and 3.57 on the Indirect Test, the former group an average of 2.61 and 5.00 on the respective tests. The *quality* of the different occupations mentioned as possibilities is also significantly different for the two groups. When the occupations mentioned were divided into conventional (lawyer, doctor, professor) and unconventional (adventurer, inventor, writer) categories, it was found that the high creatives mentioned a significantly greater proportion of unconventional occupations than did the high IQ's. For example, on the Direct Sentence Completion Test, only 18 percent of the high IQ's mentioned any unconventional occupation, whereas 62 percent of the high creatives mentioned at least one such occupation as a possibility. (See Table 3.)

The findings of the reports of the adolescents are supported by the data from the Parent Questionnaire. Parents of 24 of the high-IQ children and 24 of the high-creativity children replied to the questionnaire. Parents of 16 high-IQ and 20 high-creativity children reported that their children had expressed an interest in some career. What is most noteworthy in these reports is that 12 (75 percent) of the parents of high-IQ children mentioned occupations in but five career categories: engineering and architecture, science, medicine, law, and teaching; only 7 (35 percent) of the parents of high-creativity children mentioned career choices exclusively in these five categories. Instead, the parents of high-creativity children tended to mention such rather unconventional occupations as "veterinary" or "entertainer," such unresolved combinations as "law or music," "teaching or art," and such expressive

TABLE 3

Number and Type of Occupations Mentioned by the Experimental Groups on Direct and Indirect Sentence-completion Tests

Type of sentence-completion test and group	Number of occupations mentioned			Unusual occupations	
	Total	Mean	Standard deviation	Number mentioned	Number of students mentioning
1	2	3	4	5	6
Direct:					
High-IQ ($N = 28$)	51	1.82*	1.09	6	5**
High-creative ($N = 26$)	68	2.61*	1.41	24	16**
Indirect:					
High-IQ ($N = 28$)	100	3.57**	1.81	12	10*
High-creative ($N = 26$)	130	5.00**	1.80	29	17*

t used to test differences between means in column 3.
χ^2 used to test differences between frequencies in column 6.
* Significant at 0.05 level.
** Significant at 0.01 level.

professions as "writing" or "dancing." None of the latter were mentioned as possibilities by the parents of the high-IQ children. A summary of the data is presented in table 4.

TABLE 4

Career Choices Mentioned by Parents of High-IQ and High-creative Adolescents

	Parents of—	
Career choice mentioned	High-IQ's (N = 16)	High-creatives (N = 20)
A single major profession such as engineering or architecture, law, medicine, teaching	12	7
Other professions and occupations or several career choices	4	13

$\chi^2 = 5.71$.
$p < 0.02$.

With respect to the *quantity* of occupations considered, depending on one's point of view, one might say that the high-creativity adolescents are more "diffuse" in occupational goals than the high-IQ adolescents, or that they are more able and willing to deal with a greater range of career possibilities. As for the *quality* of occupations considered, again depending on one's point of view, one might say that the high-creativity adolescents are either more "eccentric" in their occupational goals or more able and willing to deal with career risks; that is, to take liberties with accepted standards of adult success.

2. Data with respect to the second empirical question, the "success" orientation of the two groups, were obtained from the Outstanding Traits Test and provide further insight into the career aspirations of the high-IQ and high-creativity adolescents.

The high-IQ's ranked the qualities in which they would like to be outstanding in the following order: (1) character, (2) emotional stability, (3) goal-directedness, (4) creativity, (5) wide range of interests, (6) high marks, (7) IQ, and (8) sense of humor. The high creatives ranked the qualities in the following order: (1) emotional stability, (2) sense of humor, (3) character, (4.5) wide range of interests, (4.5) goal-directedness, (6) creativity, (7) high marks, and (8) IQ. Most noteworthy here was the extraordinarily high ranking given by the high-creativity group to "sense of humor," a ranking which not only distinguishes them from the high-IQ group (who ranked it last) but from all groups to which the test had been given.

The most relevant and striking differences, however, between the two groups in the context of the present inquiry are observed in the relationship of the qualities they want for themselves, the qualities they believe lead to adult success, and the qualities they believe teachers favor. For the high-IQ group, the rank-order correlation between the qualities they would like to have themselves and the qualities they believe make for success in adult life is 0.81; for the high-creativity group, it is 0.10. The rank-order correlation between the qualities the high-IQ group would like to have and the qualities they believe teachers favor is 0.67; for the high-creativity group, it is –0.25. (See table 5.)

TABLE 5

Rank Order Correlations among Subsections of the Outstanding Traits Test, for High-IQ and High-creative Students

	Students	
Components of correlation	High-IQ (N = 28)	High-creative (N = 26)
Personal traits believed "predictive of success" and "favored by teachers"	0.62	0.59
Personal traits "preferred for oneself" and "believed predictive of adult success"	.81	.10
Personal traits "preferred for oneself" and "believed favored by teachers"	.67	−.25

In effect, the high-IQ adolescent wants the qualities he believes make for adult success and qualities similar to those he believes teachers like; the high-creative adolescent favors personal qualities which have no relationship to those he believes make for adult success and are in some ways the reverse of those he believes his teachers favor. The high-creativity adolescents are thus more "rebellious" or more "autonomous" than the high-IQ adolescents with respect to adult standards of success, depending on one's point of view, and the high-IQ adolescents are more "compliant" or more "realistic." In any case, it seems clear that there is a systematic relationship between types of cognitive functioning and types of career aspiration. The high-IQ and the high-creativity groups are characterized by differences in their career aspirations: differences all the more basic since the school achievement of the two groups as measured by standardized verbal and numerical subject-matter tests was the same.

Discussion

Several existing conceptual formulations may be cited to account for the present data. Guilford's factors of convergent and divergent thinking are highly relevant.

> In tests of convergent thinking there is almost always one conclusion or answer that is regarded as unique, and thinking is to be channeled or controlled in the direction of that answer. . . . In divergent thinking, on the other hand, there is much searching about or going off in various directions. . . . Divergent thinking . . . is characterized as being less goal-bound. There is freedom to go off in different directions. . . . Rejecting the old solutions and striking out in some new direction is necessary, and the resourceful organism will probably succeed.

In this sense, the high-IQ adolescents tend to favor "convergent" modes of problem-solving, and the high-creativity adolescents tend to favor "divergent" modes of problem-solving, whether in the cognitive function represented by performance on the intelligence and creativity measures, or in the personal-social function represented by occupational and career choice.

In the context of motivational theory, Maslow's formulations of defense and growth are similarly relevant to the present issues. He writes:[1]

> Every human being has both sets of forces within him. One set clings to safety and defensiveness out of fear, tending to regress, hanging on to the past . . . afraid to take chances, afraid to jeopardize what he already has, afraid of independence, freedom, separation. The other set of forces impels him forward toward wholeness of self and uniqueness of self, toward full functioning of all his capacities, toward confidence in the face of the external world at the same time that he can accept his deepest, real, unconscious Self . . . This basic dilemma or conflict between the defensive forces and the growth trends I conceive to be existential, imbedded in the deepest nature of the human being, now and forever into the future . . .
>
> Therefore we can consider the process of healthy growth to be a never-ending series of free choice situations, confronting each individual at every point throughout his life, in which he must choose between the delights of safety and growth, dependence and independence . . . Safety has both anxieties and delights; growth has both anxieties and delights.

In this context, the high-IQ adolescent may be seen as preferring the anxieties and delights of safety, the high-creativity adolescent the anxieties and delights of growth.

It seems that the essence of the performance of the high-creativity adolescents lay in their ability to produce new forms, to risk conjoining elements that are customarily thought of as independent and dissimilar, to "go off in new directions." The creative adolescent seemed to possess the need to free himself from the usual, to diverge from the customary behavior; he seemed to enjoy the risk and uncertainty of the unknown. In contrast, the high-IQ adolescent seemed to possess to a high degree the ability and the need to focus on the usual and to be "channeled and controlled" in the direction of the right answers, the socially accepted solution. He appeared to shy away from the risk and uncertainty of the unknown and to seek the safety and security of the already known.

---

[1] Maslow, A. H. Defense and Growth, *Merrill-Palmer Quarterly,* 3:37–38, No. 1, Fall 1956.

Moreover, the differences between the high-IQ's and the high-creatives are not restricted to their intellectual performance or to their occupational choice. The data concerning both cognitively-oriented and socially-oriented behavior are of a piece; the characteristics that describe one describe the other. The high-IQ's tended to converge upon stereotyped meanings, to perceive personal success by conventional standards, to move toward the model provided by teachers, to seek out careers that conform to what is expected of then. The high-creatives tended to diverge from stereotyped meanings, to move away from the model provided by teachers, to seek out careers that do not conform to what is expected of them. It seems that the outstanding feature of all the data is the consistency of the cognitive (as defined by performance on intellective tasks) and the personal-social (as defined by occupational choice and career aspiration) aspects of behavior.

Turning to the educational implications of the research projects reported in these sections and, indeed, of the great bulk of research dealing with creativity, there seems to be little doubt as to which of these two personal orientations is more acceptable in the majority of our social institutions. Guilford, who clearly perceived this bias, says that education "has emphasized abilities in the areas of convergent thinking and evaluation, often at the expense of development in the area of divergent thinking. We have attempted to teach students how to arrive at 'correct' answers that our civilization has taught us are correct. This is convergent thinking. . . . Outside the arts we have generally discouraged the development of divergent thinking abilities, unintentionally but effectively."

Failure to distinguish between convergent and divergent talent in our schools may have serious consequences for the future of our society. Both kinds of talent are sufficiently important to warrant attention in educational theory and practice, and it is unwise to think of divergent fantasy as simply rebellious, rather than germinal, or unconventional career choice as invariably unrealistic rather than courageous.

It is hoped that a multidimensional approach to the study of giftedness along the lines attempted in these studies will bring about an increased appreciation of the motivational and cognitive variability of children.

Part V

THE GROWTH AND DECLINE OF INTELLIGENCE

11

On the Growth of Intelligence

NANCY BAYLEY

One fact about intelligence was clearly evident to all the psychologists who investigated it, from Binet on: children develop it as they grow toward maturity. Whether or not mental growth ceases when physical growth is complete, whether or not adult intelligence declines with advancing age, whether or not there is some mathematical equation that characterizes the exact form growth curves take—these and many other questions have engaged the attention of research workers at various times. No definitive answers were possible until the results from longitudinal studies became available. The University of California at Berkeley was one of the first institutions to undertake longitudinal research on intelligence. As reports based on follow-up studies of subjects who had been members of the experimental groups from birth on began to come in, at least tentative answers to many important questions about the growth of intelligence began to emerge. At the 25-year-point, Nancy Bayley summarized these findings. The paper is an abridged version of "On the Growth of Intelligence," American Psychologist, *1955,* 10, *805–818. It is reprinted with the permission of the author and the American Psychological Association.*

One of the primary objectives of the Berkeley Growth Study (*20*) has been to study the processes of the development of intelligence as measured by tests. During the 25 years since the study was started, there has been continual change in psychologists' ideas and theories about intelligence. As a result, some of the changing theories and emphases are reflected in the series of reports that have been made on the growth of intelligence in these children. What is more, by their very nature these longitudinal studies have themselves contributed something to our knowledge of, and theories about, intellectual development and functioning.

I should like here to review the data from the Berkeley Growth Study, together with some related material, as they bear on our knowledge of the growth of intelligence.

THE CONCEPT OF THE CONSTANT IQ

In early studies the intelligence quotient was found to be very practical: it served as an easily understood index of a child's relative ability, an index by which he could be compared with other children the same age or in the same grade at school. It was found that as a rule the IQ changed very little when a child was retested a week or a month or a year or two later. This gave people great confidence in the IQ's "constancy." Furthermore, there are advantages in being able to classify a child and have him remain as classified. But this very stability of scores over certain short periods of the life span led to the assumption that intelligence is a basic entity which changes only by accretions and decrements in quantity with childhood growth, adult stability, and senescent decline. Of course this is an oversimplified extreme position, though it seems to have been (and still often is) held by many people who have used "IQ Tests" in education and in practice. It is not, however, a position held for long by those who have been actively engaged in studying the nature of intelligence and its growth.

Another result of the concept of the constant IQ has been its extrapolation, both forward and backward from the school age child, to include all ages, from birth to senescence. If the IQ is constant, then we should be able to classify a child in infancy according to his intellectual potential. We can plan his education, we can make better foster home placements, we can put the feebleminded into custodial care very young, and so on.

THE SELECTION OF INFANT TESTS

Although the Binet-Simon tests and Kuhlman's American revision included items for the first year of life, little had been done on tests of infants before Gesell set up a normative study at Yale in the early 1920's. When the Berkeley Growth Study started in 1928, we searched the literature for descriptions of infant behavior that would be suitable for evaluating intellectual development during the first year. The list tentatively compiled for our mental tests was heavily loaded with items from Gesell's norms, in their first formulations as published in 1925 (*15*). Many of these items were closely similar to those listed in other sources, but Gesell had assembled an excellent set of materials on which to test these behaviors. He was also one of the few who had actually tested a fair sample of infants, thus furnishing good preliminary norms. We selected from both the test materials and test items as first described by Gesell, adding items from other sources. In many instances we found it necessary to work out our own standard procedures and criteria of success or failure.

These tentative schedules we applied to the 61 babies of the Berkeley Growth Study, each infant being brought in at monthly intervals, starting at approximately one month of age. Ratings and descriptions were made on each child's responses during the testing situation. The items finally included in the California First Year Mental Scale were selected after analysis of their adequacy according to the usual criteria. These criteria include: their occurrence in all or most of the infants; the increasing percentage of success on them with increasing age, for appropriate developmental stages; their internal consistency and correlation with the total behavioral criterion; and their apparent relevance as intellectual, or adaptive, functions.

PREDICTION FROM SCORES IN INFANCY

At the outset we had accepted the findings based on school age children, and assumed that IQ's were constant at all ages. Consequently we were amazed at the precocity of some of the babies whose mothers seemed not very bright, and embarrassed at the poor records of other babies who, by the laws of inheritance, should have done better. But we soon found that our embarrassments and amazements were alleviated with time: a slow baby would forge ahead and redeem his inheritance, a precocious infant often seemed to rest on his laurels while the others caught up with him. We were not too surprised, therefore, when the statistical treatment of the test scores revealed that there was no relation between relative performance in the first few months of life and scores earned at the end of the first year.

When the report on the mental scores of the Berkeley Growth Study children during their first three years (*7*) was published, it was met by many with scepticism. However, in spite of their failure to conform with established theory, these Berkeley children continued to develop in their own individual ways. What is more, we have corroborative evidence from the records of the children in the Guidance Study, as reported by Honzik (*18*), and more recently by Honzik, Macfarlane, and Allen (*19*). Furthermore, these irregularities in mental growth were found to occur in other than Berkeley children. Wherever careful statistics have been applied to comparisons of repeated test scores on infants and very young children the correlations between tests separated by a year or two are low. It is now well established that we cannot predict later intelligence from the scores on tests made in infancy (*1, 16*). Scores may be altered by such conditions as emotional climate, cultural milieu, and environmental deprivation, on the one hand, and by developmental changes in the nature and composition of the behaviors tested, on the other. These latter factors are the primary concern in this paper.

As the Berkeley Growth Study children grew older we continued to record their progress by successive tests at frequent intervals. We have from time to time reported the results of these tests, along with efforts to find relationships between mental growth and other factors (*6*). When the children were 8 years old a study of the individual growth records showed that only a fifth of the group had maintained any stability in their relative status over the eight-year span (*8*). Even these few had unstable Standard Scores during the first two years.

This lack of stability in infant test scores has resulted in various efforts to supplement and to correct the infant tests to make them more predictive. It has been suggested that the scales are not composed of the right kind of test items. However, efforts to devise other, more adequate scales, invariably run into the hard fact that infants exhibit a very limited range of behaviors that can be observed and recorded. The various scales of infant intelligence have a remarkable similarity of content. At first there is little to note beyond evidences of sensory functioning in reacting to appropriate stimuli. One can observe that the one-month-old looks momentarily at a dangling ring, or at a rattle or other small object. Or one can vary the source of the sharp sound that will make him start or blink. A little later the responses are evidenced in motor coordinations: the six-month-old may pick up a one-inch cube or a teaspoon placed in easy reach. There are some early evidences of adaptation to the presented stimuli, of memory from a past experience; the seven-month-old, for example, looks "aware" that a fallen toy is no longer there, and when a little older he may turn to look for it on the floor. One can note the progression of vocalizations as they become more complex and then as they are used meaningfully. There is a developing ability to discriminate differences, to be aware of new situations, to recognize differences between members of the family and strangers, and so on.

The question is: Which, if any, among these is the forerunner of later intellectual functions? Which, if any, will predict the individual differences found in school age children?

One method of testing and selecting predictive items has been to use a later (or "terminal") measure of intelligence as the criterion. Scores earned by infants or very young children on individual test items have been correlated with their later IQ's. Those items showing the highest *r*'s with the criterion have in some instances been combined into scales. Theoretically, if other items of similar nature are then devised and added, such a scale can be expanded into an adequate predictive test. This method has been tried on infant scales by L. D. Anderson (*2*) and by Nelson and Richards (*25*). They compared successes on items under one year with retest criteria at ages 2 to 5 years. Maurer (*22*) compared scores on items in tests given children at 3 to 5 years with their scores at 15 years as a criterion. More recently

Hastings compared preschool test items with 14- to 18-year scores from our studies at the Institute of Child Welfare (*17*). In all of these studies some items have proved to be better predictors than others. Hastings selected items from the Guidance Study records and validated them on the Berkeley Growth Study (19 boys and 18 girls). His predictive items were good for the boys, but not for the girls in the validating sample. These boys had a wide range of ability, and thus an unusually large *SD* of scores. Their 2-year performance on a scale made up of good predictors correlated .67 with scores at 16–17 years. The same comparisons for the girls gave an *r* of .34. In general the predictions were better at the later preschool ages.

We have tried to find predictive items from the First Year Scale on the Berkeley Growth Study children. Several years ago, using the six children at each extreme of intelligence as measured at the 14 to 16 year tests, we went through the First Year Scale item by item, noting the age at which each of these 12 children first passed each item. We were able to select 31 items in which the six high-scoring teenagers had, as infants, been two months or more advanced over the six low-scorers. These items were an odd assortment, and there was no evident reason for their superiority over other items. Most of the items occur in the second half year, where there is a fair amount of range in scores. In the first few months very few items had a range of more than two months in age at first passing.

Recently we computed scores for the total Berkeley Growth Study sample on this 31-item scale for three ages: months 6, 9, and 12. The *r*'s of these new point scores with the mean of the intelligence sigma scores at ages 16, 17 and 18 years (for 45 cases) are .09 at six months, .32 at 9 months, and .30 at 12 months. We were unable to get significant correlations even though our sample was composed in large part of the cases on whom the items were selected, including all of the extreme cases that would determine a relationship.

So far, none of these efforts has been successful in devising an intelligence scale applicable to children under two years that will predict their later performance. The moderate successes of Maurer and of Hastings have been on items at the two-year level of difficulty or older. Even here the *r*'s are not high enough for accurate prediction on individual children. As far as I know, no one has used these items to set up and standardize an expanded scale. There does seem to be some coherence in the types of function tested by the predictive items. It is interesting to note, too, that those items which are good predictors are often not the items that best characterize a child's current stage of development. It has even been suggested that a scale should combine both types of items and then be scored in two ways—one score for evaluating present status and one for predicting future development.

These findings give little hope of ever being able to measure a stable and predictable intellectual factor in the very young. I am inclined to think that the major reason for this failure rests in the nature of intelligence itself. I see no reason why we should continue to think of intelligence as an integrated (or simple) entity or capacity which grows throughout childhood by steady accretions.

THE CHANGING ORGANIZATION OF INTELLECTUAL PROCESSES

Intelligence appears to me, rather, to be a dynamic succession of developing functions, with the more advanced and complex functions in the hierarchy depending on the prior maturing of earlier simpler ones (given, of course, normal conditions of care). The neonate who is precocious in the developing of the simpler abilities, such as auditory acuity or pupillary reflexes, has an advantage in the slightly more complex behaviors, such as (say) turning toward a sound, or fixating an object held before his eyes. But these more complex acts also involve other functions, such as neuro-muscular coordinations, in which he may not be precocious. The bright one-month-old may be sufficiently slow in developing these later more complex functions so as to lose some or all of his earlier advantage. This is the kind of thing that does seem to happen. Scores on tests given a month apart are highly correlated, but the longer the time interval between these baby tests the lower the intertest correlation.

If intelligence is a complex of separately timed, developing functions, then to understand its nature we must try to analyze it into its component parts. One approach to this process has been by factor analysis. Of the two main theories resulting from factor analysis, our data would seem to fit better into some variation of a multiple-factor than a two-factor theory. Or perhaps they fit better a theory that is intermediate, somewhere between the two.

The program of the Berkeley Growth Study has not been carried on in such a way as to make factor analysis on this material practicable. For one thing, the number of cases is too small for the usual factorial procedures. Also, for such a purpose one might have chosen a different or a more extensive series of tests. (As it is, the children have tolerated an amazingly large amount of testing and measuring!)

Nevertheless, some of our findings should point the way to new areas where factorial or other kinds of analysis would be fruitful. I should like to know, for example, where to look for *g* in the infant scales. One might expect *g* to be that factor on which prediction could be based. If *g* is not present at first, then when and how does it appear? Or does *g* itself change as its grows more complex? How do factor loadings distribute themselves in infant scales? Does a heavily-loaded first factor show a characteristic developmental process of change?

Richards and Nelson (*27*), using the Gesell items, at 6, 12, and 18 months, obtained two factors which they called "alertness" and "motor ability." They found age changes in communality of the tests that were in part due to restrictions in the type of items included in the scales at the older ages. This very fact reflects the relatively undifferentiated nature of behavior in the very young. It may be a mistake to try to call any infant behavior before 6 months more characteristically "mental" than, for example, motor. In spite of progressive selection of behaviors observed in intelligence tests, the evidence of a motor factor persists in the early ages of the Stanford-Binet, according to McNemar's factor analysis (*23*). These studies only scratch the surface of what needs to be done to gain real understanding of the nature of early mental processes.

If the word "intelligence" is best used as a broad general term that we apply to a great variety of mental functions, then we will want to investigate the nature of these functions, their interrelationships and the changes that take place in mental organization with growth. We should expect a given "factor" of intelligence to be more important at one stage of development than at another. As Garrett (*14*), has shown, in a summary of factor analyses, there is evidence of increasing independence of mental factors as children grow older. Does this trend continue indefinitely? Or do some of these factors become functionally reintegrated as they mature? The studies of Thurstone and others can be most valuable in yielding information on this point. Let us hope they will be continued over the entire life span, with careful attention to the problem of selecting items to test all relevant mental functions at all ages.

The very fact that the scores of mental growth in individual children tend to exhibit gradual shifts in relative status supports the theory that a changing organization of factors is in process. Something akin to *g*, or a high first-factor loading, must appear soon after the second or third year. The correlations of tests at these ages become positive with the later test scores. After 5 or 6 years children can be reliably classified into broad categories of normal, defective, and bright.[1]

PROBLEMS ENCOUNTERED IN CONSTRUCTING CURVES OF GROWTH IN INTELLIGENCE

The use of intelligence quotients, or standard scores, in studying growth changes in children is helpful in showing a child's shifts in status relative to the norms. But a child's progress, in relation to his own past, is better represented if we can use scores that measure increments or *amounts* of intelligence. Here we run into the problem of comparable units. Lacking absolute units for measuring intelligence, we must settle for some measure of greater or lesser difficulty, or degree of complexity of intellectual functioning. The first, and perhaps still most generally used unit of intelligence is mental age. Such a unit tends to force the same value on a mental age increment of (say) a month, whether it occurs at 6 months of age, at 6 years or 16 years. Thurstone (*30*), Thorndike (*28*), and others have tried by various devices to set up units that approximate equality of difficulty at all levels of complexity. This is done usually by comparing the overlapping distributions of scores earned by children of successively older ages. Such units would vary with the test and with the normative sample. In any event, they remain only approximations. When we accept and label them as such, however, they become useful in comparing age changes in ability.

---

[1] Since this paper was read, Hofstaetter has made a factor analysis using the 18-year consistency correlations of my data (*3*). He obtained three distinct factors: the first is predominant for the first two years, the second between 2 and 4 years, and the third accounts for almost the entire variance after 4 years. He names them: I, Sensory Motor Alertness, II, Persistence, and III, Manipulation of Symbols. This latter is most likely the general intellectual factor that most intelligence tests are designed to measure. These are, of course, global, or total-test characterizations, but they illustrate the complete break between the kinds of function measured in infants and in school-age children. (Hofstaetter, P. R., The changing composition of "intelligence": a study of *t*-technique. *J. genet. Psychol.*, 1954, *85*, 159–164.)

Thurstone applied his method to the Berkeley Growth Study scores on the California First Year and Preschool Scales for the first two years (*7*). We later extended the scaling through five years, and obtained the curve shown in Figure 1. This curve is positively accelerated for a few months, then settles into a consistent rapid growth for almost a year, after which there is a gradual slowing down in the rate, though growth continues to be fairly rapid. The curve makes sense in the light of ordinary observations of children's early development. It seems to be a useful approximation, even though one cannot claim absolute equivalence of difficulty of the units at different levels.

Problems of Equating Different Scales

When the children grew older, and an extension of the scale was in order, we ran into another problem. We had to find new tests, adapted to the children's increased mental capacities. At the beginning of the study there had been no well-standardized infant and preschool scales, and we had found it necessary to develop our own. However, good standard tests were available for school age children. The 1916 revision of the Stanford-Binet

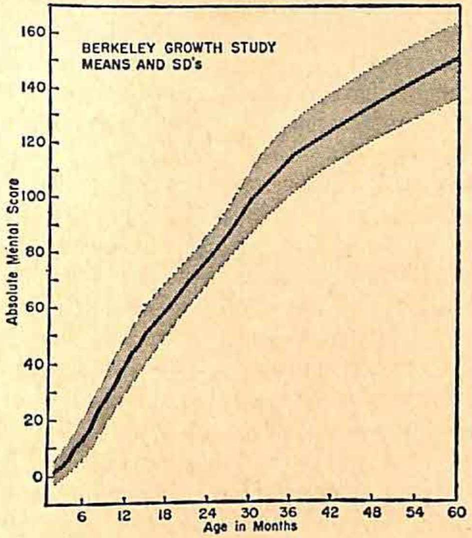


Fig. 1. Curve of intelligence, one month to 60 months, Berkeley Growth Study, according to Thurstone's method of Absolute Scaling.

was given at 6 years. Since that time we have consistently used standardized tests of intelligence, including the 1916, and both forms of the 1937 Stanford Revision of the Binet, the Terman-McNemar Group test, and the Wechsler-Bellevue Adult Intelligence Scale, Form I. But shifts in scales, with their different norms and units of increment, have complicated the problem of setting up a single continuous scale of mental growth units. This problem has not been solved, but I have approached it tentatively in several ways.

A few years ago, in presenting some data on mental growth for the first 18 years (*3*), I transposed the early scores into mental age equivalents. These scores, together with the Stanford-Binet mental ages, give us age units from birth to 17 years. The mean mental age curve of the Berkeley Growth Study children is shown in Figure 2. At 17 years, the latest available Stanford-Binet M.A. score for this group, the scores were continuing to increase. However, the rate of growth as expressed in M.A. units has diminished at the later ages. The standard deviations, shown in the lower

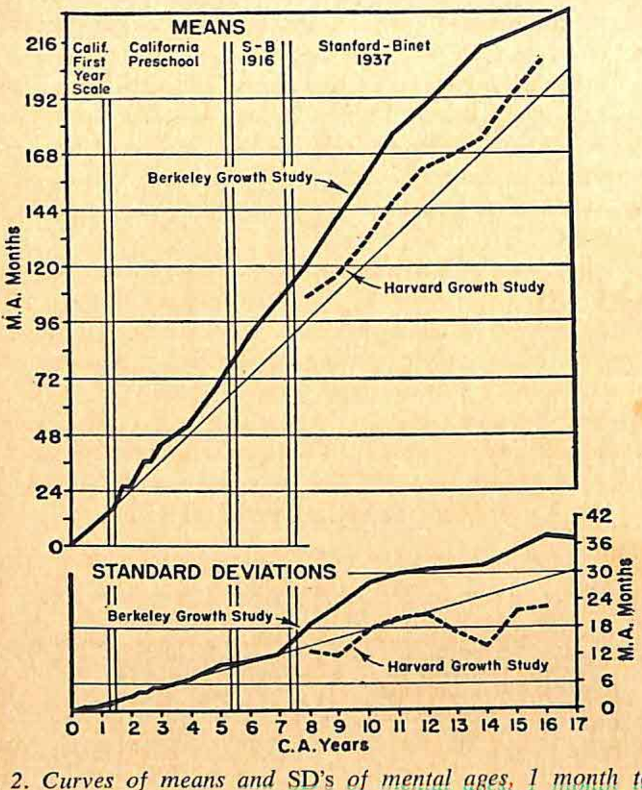


Fig. 2. Curves of means and SD's *of mental ages, 1 month to 17 years, Berkeley Growth Study. From Bayley* (3, *p.* 169).

part of Figure 2, do not increase at a constant rate but are relatively large around 10 to 12 years. Similar trends in variability may be found in other samples, and for other tests. This, I have argued (*3*), reflects a true state of increased variability in intelligence during early adolescence.

The curve of the mean mental ages is in many ways similar, for the same age intervals, to curves constructed by other investigators, using other units of mental growth. The general similarity holds whether the scores are based on longitudinal or cross-sectional samples.

Accordingly, I felt justified in using this longitudinal M.A.

curve in conjunction with the Jones and Conrad (*21*) curve based on cross-sectional data and standard score units, to construct a theoretical curve of the probable course of intelligence from birth to 60 years (*5*). . . . It is similar to curves offered by Miles (*24*), Wechsler (*31*), and others. However, more recent data on the Berkeley Growth Study, together with data from other studies, raise questions concerning the representativeness of this curve.

The data on the Berkeley Growth Study do not stop at 17 years. The Wechsler-Bellevue Adult intelligence scale was given at 16, 18, and 21 years, and we are currently repeating the test at 25 years. A study of the growth of intelligence between 16 and 21 years, as measured by these tests, is now in press (*4*). The scores were found to increase through 21 years. This was true for each category, at least to 18 years, and for the Efficiency Quotients based on the total test. It was even true of the IQ's. These trends are shown clearly in the curves of the means. . . .

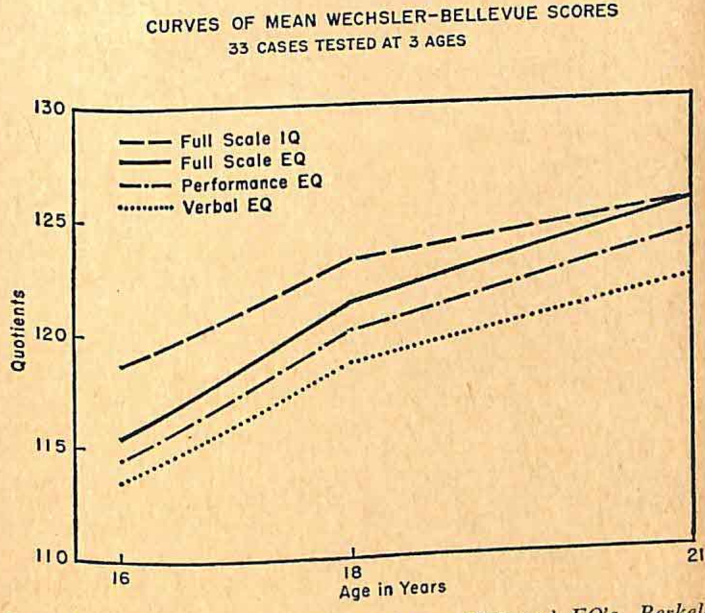


Fig. 3. Age curves of Wechsler-Bellevue IQ's and EQ's, Berkeley Growth Study. From Bayley (4).

The total scores in Figure 3 are expressed as quotients: total IQ, and EQ, and Verbal and Performance EQ. (The EQ expresses the deviation of the weighted score from the norm for 20–24-year-olds [*31*].) The gains occurred at all levels of ability within the group. All but one of the 33 subjects made some gain in total weighted score over the five-year period.

These data are in agreement with other investigators' findings on retests of the same individuals. Freeman and Flory (*13*), and Thorndike (*29*) have found for different samples, and for different tests, that intelligence scores continue to increase at least to 21 years. Dearborn and Rothney (*12*) have fitted the Harvard Growth Study data to a curve that, by extrapolation, indicates mental growth would continue to 30 years.

The general appearance of the Wechsler-Bellevue weighted score curve for the Berkeley Growth Study gives the impression that it could very well fit on as a continuation of the 17-year mental age curve. But to put the two curves into a single continuum would require transposing the scores into comparable units. This I have attempted to do, in the hope that it will give at least a rough approximation of the direction of mental growth.

The Construction of the 16D Scale

The Berkeley Growth Study tests were scheduled so that alternating forms of the Stanford-Binet were given annually through 12 years and again at 14 and 17 years. The Terman-McNemar Group test was given individually, Form C at 13 years and Form D at 15. The Wechsler-Bellevue was thus dovetailed in, having been given at years 16, 18, and 21. If we assume that the 16-year Wechsler-Bellevue scores earned by these subjects are equivalent in difficulty to a mental age at the 16-year point on their Stanford-Binet mental age curve, we can start at 16 years as a basis for equating the two sets of scores. At the adjacent ages the Stanford-Binet M.A.'s have standard deviations averaging 34 points, while the Wechsler-Bellevue *SD*'s were about 20 and the Terman-McNemar *SD*'s about 19 points.

With these data, starting with the Means and *SD*'s of the 16-year scores, we have constructed a method of transposing the scores from all tests into what may be called 16*D* Scores. That is, each child's scores at all ages are expressed in terms of the 16-year standard deviations from the mean scores at 16 years.

To do this, a table of equivalents was made by extrapolating the Terman-McNemar scores and interpolating the Stanford-Binet scores, to obtain a 16-year mean, or assumed mean, score for this sample on all three tests. To get comparable standard deviations for the three tests, the Stanford-Binet M.A. units were reduced by the fraction 20/34, and the Terman-McNemar units were increased by the fraction 20/19. Then, taking an arbitrary score of 140 to represent the 16-year mean, the three scales were related to this new 16*D* scale, point for point so that the 16-year mean would equal 140, while plus and minus one *SD* at 16 years would equal 160 and 120 respectively. This scale can be extended in either direction. All scores for all ages are expressed as deviations from the 16-year level.

The resulting curve, based on the means and *SD*'s of the 16*D* scores, is shown in Figure 4. Whether or not it is a legitimate construct, it looks reasonable, and not too far from the probable trends of growth in intelligence. It would have been better to construct the curve on a more nearly average sample, but at least we may be justified in using the 16*D* scores to apply to the members of the Berkeley Growth Study, as one way of expressing

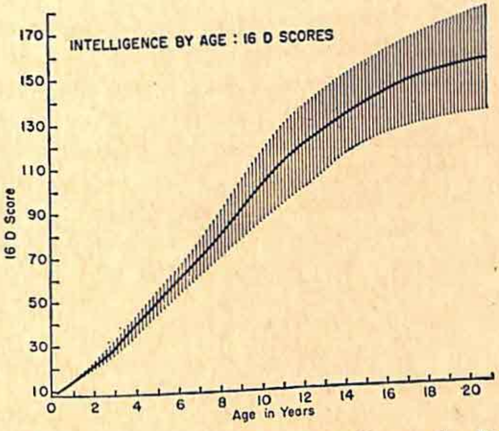


Fig. 4. Curves of means and SD's *of Intelligeenc by 16* D *units, birth to 21 years, Berkeley Growth Study.*

their progress toward, and development beyond, their status at 16 years.

Sixteen years has no particular significance as a point of reference: it was chosen because it was the only age at which the three scales we had used could be approximately equated for this sample. It would have been more satisfying if we could have started at a terminal point, say, conception, or birth, or the age at which scores stop increasing. However, 16 years is one age that has been considered a terminal point, or at least the age beyond which the ratio IQ cannot be used without modification. The score of 140 at 16 years was chosen because, from this figure when the curve is extrapolated downward from the mean score obtained at month one, the curve approximates zero at conception.

THE VALUES AND LIMITATIONS OF STANDARD SCORES AND INCREMENT SCORES

Standard Scores: Individual Curves

Let us consider some individual curves of intelligence scores earned by these subjects. In the past we have usually presented individual records in the form of Standard Scores or Sigma Scores for this group. Such scores are very useful for observing a child's changes in performance relative to others his age. We can see his ups and downs, and try to relate them to variable factors, environmental or other, that might have caused the changes. . . . We have found that the Standard Scores gave a truer picture of a child's relative status at successive ages, because there were age changes in variability of the IQ's. That is, the variability of the M.A. did not, as had been assumed, increase with age in such a way as to maintain a constant *SD* of 15 or 16 points in the MA/CA ratio. . . .

Intercorrelations

In our comparisons with such things as emotional and environmental factors that could affect test scores, we have found the

Standard Scores to be of value. For example, we have correlated the children's Standard Scores on intelligence at successive ages with the amount of schooling achieved by their parents. The age-changes in correlation (as expressed in *Z* scores) for this comparison are shown in Figure 5. The infants' scores at first are

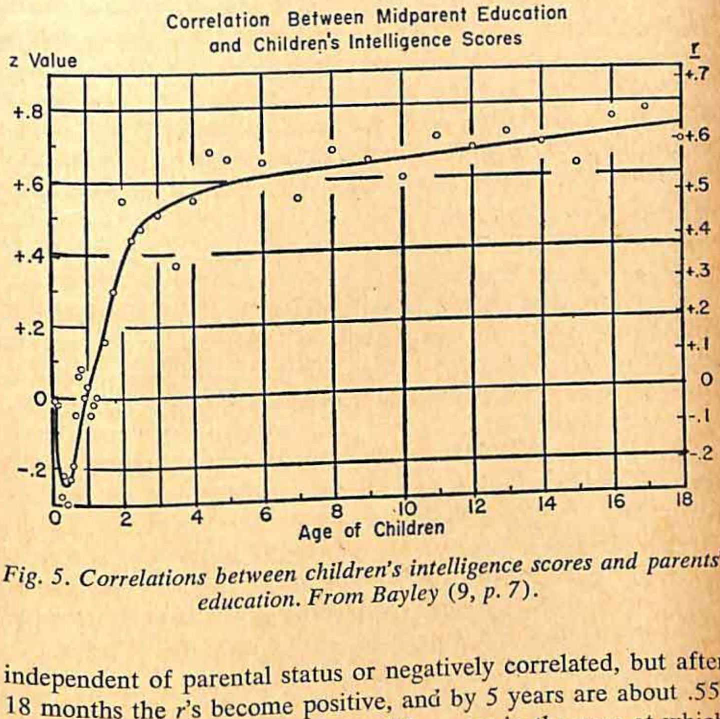


Fig. 5. Correlations between children's intelligence scores and parents' education. From Bayley (9, p. 7).

independent of parental status or negatively correlated, but after 18 months the *r*'s become positive, and by 5 years are about .55. Individual curves . . . illustrate differences in the ages at which children's scores approach the level of their parents' educational status as expressed in standard scores (*9*).

Standard scores have been used to correlate mental ability with emotional factors. For example the *r*'s between children's standard scores and the amount of time they spent crying during the period of observation and measurement were at the zero level during the first year. Then, too, the repeated standard scores obtained for one child on intelligence, can be correlated with repeated scores on other variables, using the repeated observations on a single child as a population. For example, I obtained an "Optimal" score for each testing by combining 8 ratings that were indicative of the babies' responsiveness, or attitudes that might affect their performance on the tests (*6*). The *r*'s between these Optimal scores and intelligence at any one age were close to .30. Twenty of the children had Optimal scores available for from 12 to 15 test ages each, for the age-span between 6 months and 3 years. Using the rank difference method of correlation, rho's were computed for each child between his mental standard scores and his corresponding Optimal scores. These rho's ranged from plus .77 to minus .33. For similarly-constructed "Attitude" scores, based on ratings made between 2 and 7 years of age, the individual children's rho's ranged from plus .76 to minus .46.

The wide range of correlations obtained corroborates the impression that observable emotional factors and attitudes (seen also in age curves of the different variables), rated at the time of the test, are to some extent related to the test scores, and evidently serve to help or to hinder the child's intellectual functioning. But other factors are also operative in determining a child's shifts in scores. These other factors may, in some cases, be so strong as to override the effects of emotional attitudes, resulting in negative correlations between mental performance and the child's observed responsiveness to the testing situation.

It becomes evident that the intellectual growth of any given child is a resultant of varied and complex factors. These will include his inherent capacities for growth, both in amount and in rate of progress. They will include the emotional climate in which he grows: whether he is encouraged or discouraged, whether his drive (or ego-involvement) is strong in intellectual thought processes, or is directed toward other aspects of his life-field. And they will include the material environment in which he grows: the opportunities for experience and for learning, and the extent to which these opportunities are continuously geared to his capacity to respond and to make use of them. Evidently all of these things are influential, in varying amounts for different individuals and for different stages in their growth. Many of these factors can be studied by observing concomitant variations in Standard Scores.

INDIVIDUAL DIFFERENCES IN GROWTH RATES

But Standard Scores, and other measures of relative status, have limited usefulness in the study of individual differences in rates of growth. Relative scores tend to make us forget that intellectual growth is a dynamic ongoing process, in which both averages and standard deviations in scores are related to the age of the subjects. It is worthwhile, therefore, to try to present individual curves of growth in units that will emphasize a child's change in relation to himself. Growth curves will enable us to observe a child's periods of fast and slow progress, his spurts and plateaus, and even regressions, in relation to his own past and future.

Such a growth curve has been shown in Figure 1, based on absolute scale units, for the first five years. . . . Study of the complete sample of individual curves reveals a great variety. There may be plateaus, periods of no growth, and occasionally actual decrements. There may be rapid forging ahead. Each child appears to develop at a rate that is unique for him.

By using the 16*D* scale we are now able to construct individual curves that extend for the entire period of the study. Figure 6 gives 16*D* curves for 5 boys. They cover the age span from one month to twenty-five years. Compared to the later years, the first five years seem very reduced in scope and the curves appear very homogeneous. If they are expanded to the same scale as the Thurstone curves, we find that for any given child both curves show the same periods of acceleration and retardation. The slopes of the Thurstone and the 16*D* curves are somewhat different, but the patterns of accelerations and retardations are generally similar in nature.

Although each child has his own individual pattern of progress, the patterns are not completely random. After the period of infancy there is a strong underlying consistency or constancy. Some children forge ahead and maintain relatively advanced positions after 5 or 6 years of age. Others grow slowly and lag behind. There is some shifting of position, but the changes are gradual over rather long intervals of time. Within such intervals we can expect to obtain fairly constant Standard Scores (IQ's).

It is notable that these five boys have all been tested at 25

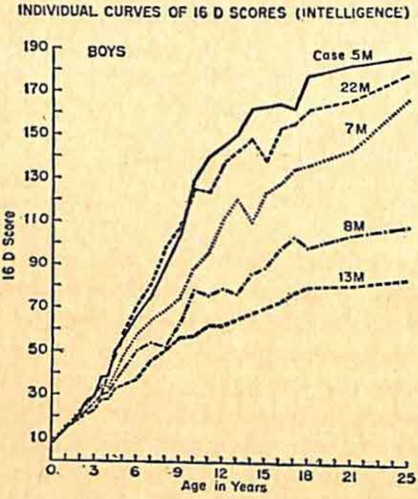


Fig. 6. Individual curves of intelligence (16 D *units) for 5 boys, one month to 25 years. Berkeley Growth Study.*

years, and all five have continued to improve in their Wechsler-Bellevue scores. The continued growth occurs at all levels of ability. Case 13M, the slowest boy, has had increments in his Wechsler-Bellevue IQ's from 63 at 16 years to 78 at 25 years. This boy spent much of his childhood (ages 10 to 23 years) in an institution for the mentally retarded. When tested at 21 years he had never learned to read more than a few words. Now at 25 he reads, slowly to be sure, but he read aloud without error the Wechsler-Bellevue arithmetic problems. . . .

Some of the dips in the individual curves are due to changes in the tests. For example, those who have trouble in reading make relatively low scores on the Terman-McNemar Group test. But often the irregularities cannot be attributed to changes in the tests used at different ages.

Slight irregularities may reflect temporary conditions of motivation, health, or emotional factors. The more constant shifts require other explanations. Though they may result from prolonged emotional or environmental influences, they may also express inherent tendencies to develop at given rates. I suspect that each child is a law unto himself: in some instances certain factors are more important, while in others different factors play the determining role.

TEMPORAL CHANGES IN ADULT INTELLIGENCE

The few 25-year scores so far available indicate that the intellectual processes measured by these tests have not yet reached a ceiling. Fourteen out of fifteen subjects tested show continued increments. If these are typical cases, what, then, may we venture to predict for the years ahead? The alternative explanation of practice effects from repeating the same test might be offered. But the intervals between repeats on the Wechsler-Bellevue are 2, 3, and 4 years. These are rather long times to remember much about the specific items. Nevertheless, there is probably some residual memory for, or vague familiarity with, the task and the type of solution found at the previous testings. At present we must assume that these factors account for part of the increment.

On the other hand, we have some recent evidence that some intellectual functions do continue to improve with age in adults. When the same individuals are retested after long intervals on the same test or on an alternate form of a test, the scores on the retests are significantly higher. These retests were carried out on superior adults, and their patterns of mental change may be different from those of less able persons.

In a recent study of the adult intelligence of the subjects of the Terman Study of Gifted Children, Bayley and Oden (*10*) found that scores on the difficult Concept Mastery test increased on a second testing. For a population of over a thousand, composed of Gifted Study subjects and their spouses, comparisons were made between two tests that had been taken about 12 years apart. The increase in scores on the retest averaged about half a standard deviation. The subjects ranged in age from about 20 to about 50 years. When they were grouped into 5-year age intervals, the test-retest scores of all age groups increased. . . .

Similar results have been reported by Owens (*26*) who repeated the Army Alpha test at 50 years on 127 men who had first taken the test as 19-year-old freshmen at Iowa State College. Their scores improved by .55 *SD*'s over the 31-year interval. One can hardly claim practice effects after a lapse of 31 years. Even the 12-year interval of the Terman study is rather long for any such claim: also the Gifted Study subjects were retested on an alternate form, thus ruling out specific memories of items. Furthermore, there were control groups consisting of those who were tested only once, at either the 1940 or the 1951 testing. The differences in mean scores of these groups at the two testings are the same as for the twice-tested groups.

A SUGGESTED FIFTY-YEAR CURVE OF INTELLIGENCE

I have experimented with using the data from these two studies of adults to extend the 16*D* growth curve to 50 years. The subjects of the Berkeley Growth Study are, on the average, a somewhat superior group. Their 16-year Wechsler-Bellevue mean IQ is 117, and their 17-year Stanford-Binet mean IQ is 129. A small group of 25-year-olds who have taken the Concept Mastery earned scores close to the average for the spouses of the Terman subjects at that age. We may assume, then, that this sample is rather similar to the Iowa State Freshmen and to the spouses of the Gifted Study subjects, in its general level of test performance. It has, therefore, seemed reasonable to join the data from the Berkeley Growth Study directly to the scores of either of the other studies, in extending the curve, as in Figure 7.

This joining of the curves has been done for the Iowa study simply by placing the 19-year initial point at 19 years on the 16*D* curve and the 50-year point at the equivalent on the 16*D* scale of an increase of .55 standard deviations.

For the Gifted Study spouses the process was a little more complicated, but it has yielded a series of intermediate points, giving some indication of the probable shape of the curve. To obtain these points, I plotted a series of *SD* increment curves, placing the successively older ages at points on the curves of the younger groups in such a way as to take into account the growth

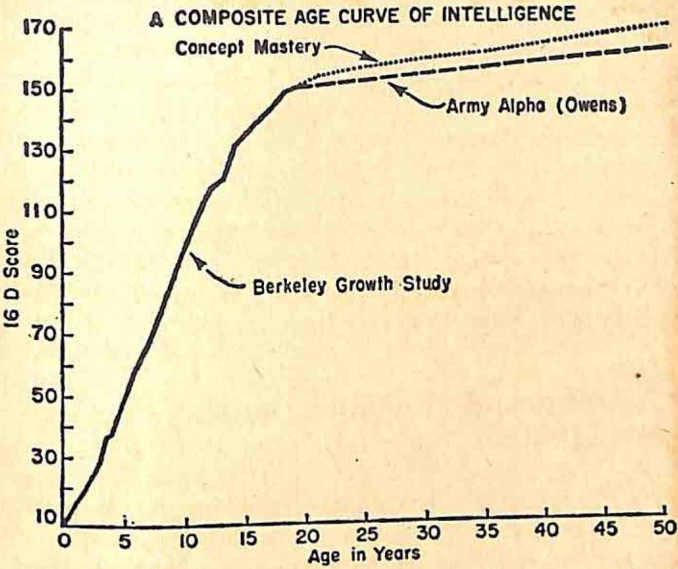


Fig. 7. A proposed age curve of intelligence, birth to fifty years. Based on data from Berkeley Growth Study, the Terman Gifted Study and Owens Iowa Study.

already attained at any new starting age.* From these series of overlapping curves, a smoothed curve was drawn, and equivalent 16*D* scores were read off at 5-year intervals.

The resulting two-pronged curve for the 50-year span shows a more modest increment for the Alpha scores of the Iowa men. The Concept Mastery scores of Gifted Study spouses gain a full standard deviation, or about twice as much. Of course, since both of these curves are only approximations, neither may be more correct than the other. The differences are probably due, at least in part, to differences in the testing instrument. The Concept Mastery scale for one thing, has far more top than the Alpha, and allows for much greater expansion upward.

---

* That is, the youngest group was tested at 20 years and again at 32, with a gain of .4 *SD*. The 25-year initial score of the next older group was then plotted at the 25-year point on the first curve (or at .19 *SD*) and their gain at 37 years was plotted as .33 *SD* above this point (or at .52 *SD*), and so on for successively older groups.

We have here evidence that tested intelligence, as measured by verbal concepts and abstractions, continues to grow when populations composed primarily of superior adults are retested. Intelligence may also continue to increase in the less bright. Certainly, the less favored members of the Berkeley Growth Study are still improving their scores at 25 years. What is more, in several other studies there is evidence that this phenomenon is not confined to individuals tested at the University of California Institute of Child Welfare. Freeman and Flory (*13*), for example, divided the children in their study on the basis of scores at 12, 13, and 14, into low and high scores. At the later ages, 16 and 17 years, the low-scoring group was continuing to improve at a faster rate than the high-scoring group. A recent study by Charles (*11*) reports retest IQ's for 20 adults who had been diagnosed in childhood as feeble-minded. Their mean childhood Stanford-Binet, 1916, IQ was 58 and their mean adult Wechsler-Bellevue IQ was 81. Charles accounts for this difference in two ways: errors of diagnosis in childhood, and evidence from other studies that people who score low on the Binet test tend to make higher scores on the Wechsler. Similar explanations have been offered for similar findings in other studies. But a mean increase of 23 IQ points amounts to 1.5 *SD*'s of either of the tests used. This is a rather large shift to be attributed to test differences in restriction of scores, to regression phenomena, or to errors in the original test. All 20 individuals improved on the retest. It seems to me quite possible that these people did continue to improve in their mental ability.

There are many gaps in our knowledge of the nature of intelligence, and many questions remain unanswered concerning age changes in mental organization. In the curve presented in Figure 7, there remains an unanswered discrepancy between the adult portion and data for these ages presented by earlier investigators, who have found decrements in scores with increasing age after about 21 years. In the earlier studies some types of functions held up better than others. Owens found that those abilities that had held up best on the cross-sectional samples were the same ones that increased the most on his retests. The real difference between the conflicting findings seems to lie in the longitudinal as opposed to the cross-sectional method of obtaining scores for successive ages. In the former we have a constant sample whose life experiences, age for age, will have been similar in pervasive environmental conditions, such as wars, technological advances, and methods of education.

If, after taking adequate account of practice effects, the increases still remain, then the next question is to inquire into the nature of the tests, and the extent to which they measure intellectual abilities. Do such tests as the Army Alpha and the Wechsler-Bellevue, for example, measure intelligence in adults? Or do they tend to reflect continued experience in an increasingly enriched environment? Do the younger generations have more opportunity to develop their intellectual capacities than did their parents, or even their older brothers and sisters? Or are we just measuring the effects of increasingly widespread informal education made possible by radio, television, and other modern means of communication?

If, regardless of the cause of the improved scores, they reflect actual degrees of competence outside of the testing situation, then these scores continue to have practical value. Another practical question is: What norms should be used in measuring deterioration resulting from brain injury, or from senescence? Perhaps it will be necessary to compare a present 50-year-old man's score with norms for, say, those who are 50 in 1954, rather than with 50-year norms for other decades.

What normal age changes should we expect in mental organization? The curve presented here is a composite. The forms of growth curves vary according to the functions measured. We should expect differences in the steepness of increment and decrement in growth curves of the different functions, and differences in the ages at highest efficiency. These differences have been found consistently in cross-sectional studies. The question raised here is whether more adequate studies, of the same individuals through time, will not show that the age of highest intellectual capacity is later than we thought, and that the decrements in abilities are, correspondingly, deferred.

This curve is offered as an alternative to previously published age-curves of intelligence. I should like to see it tested with further research that would refine, modify, and extend it into a more complete and accurate representation of intellectual changes over the entire life span.

REFERENCES

1. ANDERSON, J. E. The limitations of infant and preschool tests in the measurement of intelligence. *J. Psychol.*, 1939, 8, 351–379.
2. ANDERSON, L. D. The predictive value of infancy tests in relation to intelligence at five years. *Child Develpm.*, 1939, 10, 203–212.
3. BAYLEY, NANCY. Consistency and variability in the growth of intelligence from birth to eighteen years. *J. genet. Psychol.*, 1949, 75, 165–196.
4. BAYLEY, NANCY. Data on the growth of intelligence between 16 and 21 years as measured by the Wechsler-Bellevue Scale. *J. genet. Psychol.*, in press.
5. BAYLEY, NANCY. Development and maturation. In H. Helson (Ed.), *Theoretical foundations of psychology*. New York: D. Van Nostrand, 1951. Ch. 4, pp. 145–199.
6. BAYLEY, NANCY. Factors influencing the growth of intelligence in young children. *Yearb. nat. Soc. Stud. Educ.*, 1940, 39, 49–79.
7. BAYLEY, NANCY. Mental growth during the first three years. *Genet. Psychol. Monogr.*, 1933, 14, 92.
8. BAYLEY, NANCY. Mental growth in young children. *Yearb. nat. Soc. Stud. Educ.*, 1940, 39, 11–47.
9. BAYLEY, NANCY. Some increasing parent-child similarities during the growth of children. *J. educ. Psychol.*, 1954, 45, 1–21.
10. BAYLEY, NANCY, & ODEN, M. H. The maintenance of intellectual ability in gifted adults. *J. Gerontol.*, 1955, 10, 91–107.
11. CHARLES, D. C. Ability and accomplishment of persons earlier judged mentally deficient. *Genet. Psychol. Monogr.*, 1953, 47, 3–71.
12. DEARBORN, W. F., & ROTHNEY, J. W. M. *Predicting the child's development*. Cambridge: Sci-Art, 1941.
13. FREEMAN, F. N., & FLORY, C. D. Growth in intellectual ability as measured by repeated tests. *Monogr. Soc. Res. Child Develpm.*, 1937, 2, 116.
14. GARRETT, H. E. A developmental theory of intelligence. *Amer. Psychologist*, 1946, 1, 372–378.
15. GESELL, A. *The mental growth of the pre-school child*. New York: Macmillan, 1925. Pp. 447.
16. GOODENOUGH, FLORENCE L. *Mental testing*. New York: Rinehart, 1949. Pp. 609.

17. HASTINGS, HOMER. The predictive value of individual items in preschool intelligence tests. Unpublished doctor's dissertation, Univer. of California, 1952.
18. HONZIK, M. P. The constancy of mental test performance during the preschool period. *J. genet. Psychol.*, 1938, 52, 285–302.
19. HONZIK, M. P., MACFARLANE, JEAN W., & ALLEN, L. The stability of mental test performance between two and eighteen years. *J. exper. Educ.*, 1948, 17, 309–324.
20. JONES, H. E., & BAYLEY, NANCY. The Berkeley growth study. *Child Develpm.*, 1941, 12, 167–173.
21. JONES, H. E., & CONRAD, H. S. The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty. *Genet. Psychol. Monogr.*, 1933, 13, 223–294.
22. MAURER, K. M. *Intellectual status at maturity as a criterion for selecting items in preschool tests.* Minneapolis: Univer. of Minnesota Press, 1946.
23. MCNEMAR, Q. *The revision of the Stanford-Binet Scale: An analysis of the standardization data.* Boston: Houghton Mifflin, 1942.
24. MILES, R. Psychological aspects of ageing. In E. V. Cowdry (Ed.), *Problems of ageing* (2nd Ed.). Baltimore: Williams & Wilkins, 1942. Pp. 756–784.
25. NELSON, V. L., & RICHARDS, T. W. Fels mental age values for Gesell schedules. *Child Develpm.*, 1940, 11, 153–157.
26. OWENS, W. A. Age and mental abilities: A longitudinal study. *Genet. Psychol. Monogr.*, 1953, 48, 3–54.
27. RICHARDS, T. W., & NELSON, V. L. Abilities of infants during the first eighteen months. *J. Genet. Psychol.*, 1939, 55, 299–318.
28. THORNDIKE, E. L. *The measurement of intelligence.* New York: Teachers College, Columbia Univer., 1926.
29. THORNDIKE, R. L. Growth of intelligence during adolescence. *J. genet. Psychol.*, 1948, 72, 11–15.
30. THURSTONE, L. L. A method of scaling psychological and educational tests. *J. educ. Psychol.*, 1925, 16, 433–451.
31. WECHSLER, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.

Part VI

THE NATURE-NURTURE CONTROVERSY

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

12

Schooling Makes a Difference

IRVING LORGE

The issue around which the most acrimonious debates have raged whenever intelligence testing is under consideration is whether the individual differences such tests reveal arise mainly from heredity or mainly from environmental influences. Psychologists, sociologists, educators, and many other varieties of scholars and professional men were concerned with this question because of what they thought to be its implications for social policy. As time has passed, the progress of research has served to moderate extreme views and to show how the concepts stressed by both hereditarians and environmentalists could be merged in more comprehensive theories of human development. To both questions intelligence testers have tried to answer—"Do differences in score reflect hereditary differences in individuals?" and, "Do differences in score arise from differing environments?"—the answer is "Yes."

Out of all the environmental differences we observe in human society, the one variety that influences mental growth most markedly is differences in education. The evidence for this statement comes primarily from retests of equated groups of subjects after considerable periods of time. Invariably the trend is for those who have had more schooling to score significantly higher than those who have had less schooling. An influential report of such a retest investigation by Irving Lorge is reprinted here, "Schooling Makes a Difference," Teachers College Record, *1945, 46, 483–492. It is reprinted with their permission.*

Educators recognize the value of education for increased earning power. They also recognize the value of education for increased knowledge. But educators do not realize the value of education for increased mastery of the abilities measured by intelligence tests.

Few studies have shown the influence of extent of education upon changes in intelligence test scores. Except for the Iowa Studies,[1] none has shown how increased educational advantages can influence the scores made on intelligence tests. This article is a report of the changes which occurred in intelligence test scores as related to extent of education for a group of boys who were studied over a twenty-year period—from the time the boys were about fourteen years old to the time they were about thirty-four years old.

In 1921–22, some 863 boys in the 8B classes of a representative sampling of public elementary schools in the city of New York were given a series of tests of abstract intelligence, mechanical adroitness, and clerical ability. The tests were given to ascertain the values of such appraisals in predicting subsequent vocational success. The boys were followed up at regular intervals to get relevant information about their subsequent schooling and subsequent work careers.[2]

RETESTS AFTER TWENTY YEARS

In May 1941, as the twentieth year of follow-up was nearing its close, the 863 boys were asked to come to Teachers College for some retesting.

The primary objective of the retest program was to find out whether an intelligence test taken as a thirty-four-year-old is more predictive of vocational success than an intelligence test taken as a fourteen-year-old. The letter sent to these subjects stated that twenty years had passed since they had been tested and asked them to come to the College to meet with their elementary school classmates, and to take some further tests. They were told to bring their work records up to date, since these records had been of great value earlier in appraising aspects of vocational guidance, and each bit of additional information would increase their value. Each man was offered (and paid) two dollars for his time. During May 1941, 131 men complied with the request.

---

[1] Guy Montrose Whipple (Editor), "Intelligence: Its Nature and Nurture, Part I," *The Thirty-Ninth Yearbook of the National Society for the Study of Education* (Chap. 14, pp. 405–41). Public School Publishing Company, Bloomington, Ill., 1940.

[2] Edward L. Thorndike and others, *Prediction of Vocational Success,* Commonwealth Press, New York, 1934.

Each of the 131 men was interviewed to complete his work record for the period that had elapsed since he was last seen. Each was asked to take the Minnesota Scale for the Survey of Opinions and to fill out a questionnaire about his education, social affiliations, ambitions, and opinions concerning guidance. In addition, each took two tests of intelligence: the Otis Self-Administering Test of Mental Ability, Higher Examination, Form B (twenty-minute time limit), and Part III of the Thorndike Intelligence Examination for High School Graduates, Form V (one-hour time limit). The present article is concerned with the relationship of these tests of intelligence to those taken twenty years earlier.*

A REPRESENTATIVE SAMPLING

On the basis of the scores obtained in the 1921–22 testing there was no statistically significant difference on any test between the average score of the group of 863 boys and the average score of the 131 men who took the 1941 retests. Furthermore, there was no statistically significant difference between the measure of variability of the 863 boys and the 131 men. An additional check was made to determine whether any statistically significant difference existed between the intercorrelations among the six measures obtained in 1921–22 for all 863 boys and the corresponding intercorrelations for the 131 men of the retest group. There were none. It is reasonable, therefore, to infer that the 131 men who came to Teachers College in May 1941 constituted a representative sampling of the total group tested twenty years before. Since the original group of 863 boys constituted a representative sampling of boys in grade 8B in public elementary schools in New York City in 1921–22, the retest sample may be considered representative of that group too.

---

* For each of the 131 men (and, of course, for each of the original group of 863 boys), records of the results of the 1921–22 examinations and of certain information from his elementary school career were available. The 1921–22 data included the following facts: age, scores on two clerical abilities tests, the Stenquist Mechanical Assembly Test, a group test of intelligence, and a measure of school progress up to grade 8B (the second half of the eighth grade). It was possible, therefore, to estimate whether the 131 men were a representative sample of the entire group of 863 boys originally tested.

From a study of the relationships among the 1921–22 test of intelligence, the 1941 tests of intelligence, and the extent of schooling for the 131 men, it is possible to make certain generalizations concerning the population of boys in the eighth grade in 1921–22. The score on the 1921 test of intelligence was a weighted composite of the scores from two different tests: the Thorndike-McCall Reading Scale and the I.E.R. Arithmetic Test. Toops[3] has shown that this composite score derived from a reading and an arithmetic test measures essentially what is measured by other reputable group tests of intelligence. The 1921 score, therefore, may be considered a valid indication of how these boys performed on intelligence tests when they were nearly fourteen years old. The two tests of intelligence administered in 1941 are reputable group tests of intelligence. For the 131 men in the retest sample, the correlation between the Otis scores and the Thorndike scores is .82, suggesting that each of the tests measures similar intellectual performances.

HIGHEST GRADE COMPLETED

The extent of schooling was measured in terms of the highest grade completed in school. As such, it is comparable with the census definition of "last school grade completed." In arriving at the score for extent of schooling, the last school grade completed was credited regardless of the kind of school in which it was completed. A person who had completed the tenth grade in an academic high school was given the same score as a person who had finished the tenth grade in a vocational high school, or one who had completed the equivalent of the tenth grade in an evening school, or one who had finished the equivalent of the tenth grade in a school established by the company by which he was employed. The extent of schooling as measured by the highest grade completed in school is, therefore, an undifferentiated mixture of quantity of education, kind of education, and quality of education. It does not reveal what the individual really gained from his education, nor is it an index of the quality of education he received. The highest grade completed is reported in whole numbers; thus, a score of 8 represents the equivalent of graduation from the eighth grade of elementary school, a score of 12 graduation from high school, a score of 16 graduation from college, and a score of 17 or more is indicative of one or more years of postgraduate study.

---

[3] Herbert A. Toops, *Tests for Vocational Guidance of Children Thirteen to Sixteen.* Contributions to Education, No. 136. Bureau of Publications, Teachers College, Columbia University, New York, 1923.

INTELLIGENCE TESTS IN 1921 AND 1941

The correlation between the scores on the 1921–22 test of abstract intelligence and the Otis intelligence test taken in 1941 was .64. The correlation between the 1921–22 abstract intelligence score and the score on the Thorndike test taken in 1941 was .62. There is no doubt that this is a rather high correlation between two *different* tests of intellectual performance taken twenty years apart. Robert Thorndike[4] has shown the correlation between the scores derived from the *same* intelligence test (the Stanford Revision of the Binet-Simon Scale) taken five years apart to be about .70. His data suggest that the correlation would be substantially lower if the scores were taken ten, fifteen, or twenty years apart. There is approximately as much in common in the intellectual performance based on different tests taken twenty years apart as there would have been if the same test of intelligence had been repeated twenty years later.

---

[4] Robert L. Thorndike, "The Effect of the Interval Between Test and Retest on the Constancy of the I.Q.," *Journal of Educational Psychology*, 24:543–49, October, 1933.

THE VALUE OF SCHOOLING

In an earlier study, on the basis of the repetition of identical tests ten years apart (1921–22 and 1932) it was found that time diminishes the predictive value of a test for itself. Factors related to maturation, environment (especially school), and the test itself may be responsible for the general decrement in the correlation between scores on intelligence tests taken at widely separated times.[5] In the 1941 study, one of the environmental factors can be appraised. The value of the extent of schooling can be estimated by finding out whether the addition of the highest grade completed affects the relationship between the 1921 and the 1941 intelligence test scores. The correlations between the two tests of intelligence twenty years apart were .64 with the Otis and .62 with the Thorndike. Does highest grade completed increase the prediction? The facts suggest a positive answer. The multiple correlation between the 1921–22 test of intelligence and the composite of the Otis and the highest grade completed is .79. The corresponding correlation between the 1921–22 test of intelligence and the composite of the Thorndike and the highest grade completed is .77. The increase in the prediction from .64 to .79 and from .62 to .77 definitely suggests that extent of schooling does make a difference in the intelligence test score of a person twenty years later.

RELATIONSHIP OF ADULT I.Q. TO SCHOOLING AND CHILDHOOD I.Q.

The extent of this difference can be understood somewhat better from tables which show the relationship of the 1941 tests of intelligence to the 1921 test of intelligence and extent of schooling. Table 1 presents the average scores on the Otis test given in 1941 in terms of highest grade completed and the score on the 1921 test of abstract intelligence. The columns of the table represent the average Otis score in 1941 for the men who had specified scores on the 1921–22 test of abstract intelligence.

---

[5] Irving Lorge, "Retests After Ten Years," *Journal of Educational Psychology*, 25:136–41, February, 1934.

TABLE 1

Highest grade completed by 1941	Score on 1921–22 test of abstract intellegence											
	49–58		59–68		69–78		79–88		89–98		99–114	
	Otis Score	*N*	Otis Score	*N*	Otis Score	*N*	Otis Score	*N*	Otis Score	*N*	Otis Score	*N*
8	14.0	4	22.0	4	20.7	9	26.4	5	39.0	2	33.0	1
9	19.0	1	19.5	2	14.5	2	31.1	8	38.0	2	29.0	1
10	24.0	1	22.0	4	25.1	9	28.5	8	37.0	4	46.5	2
11 or 12	21.0	1	26.0	1	31.7	3	31.0	9	41.0	3	34.0	1
13 or 14			22.0	1	26.0	3	34.7	4	41.7	4	37.5	2
15 or 16			34.0	1	27.0	1	39.5	6	53.5	2	50.8	5
17 or more					38.0	3	46.0	5	54.5	6	43.0	1

Average scores on the Otis Self-Administering Test of Mental Ability, Higher Examination, Form B (20-minute time limit), taken in May 1941, classified by score on the 1921–22 test of abstract intelligence and by highest grade completed by May 1941: for 131 boys who may be considered a representative sample of the Vocational Guidance Grade Group.

For instance, in the column headed 49–58 are presented the average Otis scores for the seven boys who had the lowest 1921 scores. Four of these boys who did not go beyond grade eight made an average Otis score in 1941 of 14.0. One of these boys who completed the ninth grade made an Otis score of 19.0; another who completed the tenth grade made an Otis score of 24.0. Each column, therefore, shows the 1941 scores of men who may be considered to have been fairly equivalent in measured intellectual ability in 1921.

The rows of Table 1 present the average Otis score obtained in 1941 for boys who had completed grade eight, or nine, or ten, or above. For instance, in the first row (representing completion of eighth grade only) the average Otis scores are given for men whose extent of education was constant, but whose 1921–22 intelligence scores differed. Of the twenty-five men who had completed the eighth grade only, the four who had the lowest 1921–22 intelligence test score made an average of 14.0; of the same group of twenty-five, the one person who had made a score in the 99–114 range of abstract intelligence test scores in 1921–22 obtained a score of 33.0 on the Otis test in 1941.

Table 2 presents the corresponding data for the Thorndike Intelligence Examination for High School Graduates taken in 1941 in terms of the 1921–22 test scores and the highest grade completed.

The two tables, therefore, give the information regarding the relationship between the 1921–22 test of intelligence and the 1941 tests of intelligence, holding constant the extent of education (reading across the row), and the information regarding the relationship between the 1941 tests of intelligence and the extent of education, holding constant the score on the 1921–22 test of intelligence (reading down the columns).

The facts in Table 1 and Table 2 definitely indicate that there is a substantial relationship between the scores on 1921 test of intelligence and the scores on 1941 tests of intelligence, holding constant the extent of education. The men who had completed only eight years of schooling show average Otis scores of 14, 22, 21, 26, 39 and 33, depending on their 1921–22 intelligence test scores. For each row it is clearly evident that the average

Table 2

Highest grade completed by 1941	Score on 1921–22 test of abstract intelligence											
	49–58		59–68		69–78		79–88		89–98		99–114	
	Thorndike Score	*N*	Thorndike Score	*N*	Thorndike Score	*N*	Thorndike Score	*N*	Thorndike Score	*N*	Thorndike Score	*N*
8	21.5	4	34.7	4	44.5	9	51.6	5	61.0	2	49.0	1
9	40.5	1	53.0	2	42.2	2	48.8	8	74.0	2	65.0	1
10	33.5	1	45.2	4	37.6	9	53.2	8	68.1	4	76.0	2
11 or 12	24.5	1	61.0	1	51.0	3	62.2	9	87.7	3	93.5	1
13 or 14			52.5	1	67.5	3	71.0	4	77.0	4	66.5	2
15 or 16			63.0	1	60.5	1	69.9	6	93.5	2	92.6	5
17 or more					69.3	3	87.8	5	95.5	6	84.5	1

Average score on Part III of the Thorndike Intelligence Examination for High School Graduates, Form V, taken in May 1941, classified by score on the 1921–22 test of abstract intelligence and by highest grade completed by May 1941: for 131 boys who may be considered a representative sample of the Vocational Guidance Grade Group.

score on the 1941 tests of intelligence is higher for those with 1921–22 intelligence test scores in the range 99–114 than for those in the range 49–58. The data definitely suggest that a person's measure of intellectual ability taken as an adult 34 years of age is substantially related to his measure of intelligence at or near 14 years of age, regardless of the extent of his schooling.

A more important finding, however, is concerned with the variation in the 1941 measure of intellectual ability as related to the extent of education when the 1921–22 measure of intelligence is held constant. In Table 1, for instance, in the column representing the Otis scores of those who had a 1921–22 intelligence test score of 89–98, the scores tend to increase with the extent of schooling. The averages for eight, nine, and ten grades completed are 39, 38, and 37 respectively. The averages for the boys who had the same 1921–22 intelligence scores and who had fifteen or more grades of education were 53.5 and 54.5. In terms of the published Binet equivalents for the Otis scores, the mental age represented by scores of 37, 38, and 39 is approximately 16 years 5 months, whereas the mental age equivalent of scores of 53.5 and 54.5 is approximately 18 years 5 months. The mental age of the eight men in the range 89–98 with fifteen or more grades of schooling is two full years higher than that of the men who completed between eight and ten grades of schooling. If one used 16.0 as the reference chronological age to estimate I.Q.'s, the men who had completed fifteen or more grades would have I.Q.'s of about 115 as compared with an average I.Q. of 103 for those men who had completed only eight to ten years of schooling, although each of them had the same intelligence test score at fourteen years of age. The 1941 Otis scores for men who had completed thirteen or more grades of schooling are substantially higher than those for men who had completed less than twelve grades of schooling, holding constant the level of intelligence obtained in 1921–22.

The tables reveal another finding of considerable significance. It should be noted that practically none of the men who had the lowest intelligence scores in 1921–22 (49–58 and 59–68) made scores as high as the men who had obtained the highest intelligence scores in 1921–22 (89–98 and 99–114). Education makes a difference, but it would be overly optimistic to expect it to change the least able into the most able.

The men highest in 1921–22 intelligence test score are differentiated according to the extent of education they obtain. The equally able men who had less than twelve grades of education obtain adult intelligence test scores substantially lower than their intellectual peers who had more than thirteen grades of schooling. Insofar as an intelligence test measures ability to perform intelligently, the men who did not have the advantages of more schooling apparently are penalized by this lack of education. It must be recognized that superior intellectual ability needs stimulation. Intellectual ability is not lost as people grow older, but the full potentialities of a person may be lost in the absence of reasonable opportunities and advantages. Genius may come out, but society ought to insure its emergence by providing appropriate educational opportunities.

SELECTIVITY OF EDUCATION

Education has been selective of the abler people in the population. The selection, however, has not been perfect. For the 131 men involved in this study, the 1921–22 test of intelligence correlated only .36 with highest grade completed. A correlation coefficient of .36 would not permit very reliable predictions of how much schooling a person would complete. The relationship between the 1941 measure of intelligence and the extent of education is much higher. For the Otis it was .67 and for the Thorndike it was .68. An intelligence test score obtained at or near age fourteen does not predict the extent of education a person will get—and the intelligence test score obtained at or near age thirty-four does show a substantial relation to the extent of education he has received. Schooling, therefore, makes a difference in the intelligence test score that an adult will obtain. It is difficult to estimate the value of these changes in scores in measured mental ability that favor the men with more education. In terms of intellectual performances involved in dealing with ideas and comprehension of them, the differences must be significant.

THE MEANING OF INTELLIGENCE TEST SCORES

An adult's measured mental ability is related to his intelligence as a child and to the extent of his subsequent schooling. An intelligence test score, therefore, is related not only to basic mental ability but also to the environmental opportunities to capitalize upon that basic ability. Intellect—that is, native intellect—can never be measured directly. Mental ability can be appraised only by its performances. An intelligence test is made up of items or situations each of which must be an environmental task. All other things being equal, it is assumed that persons within the same environment can be differentiated according to their test performances. This assumption, undoubtedly, is not true when variation in the extent of schooling can differentiate the initially able persons' behavior on adult intelligence tests. Moreover, the test situations should ideally be a sample of all possible environmental tasks requiring intelligent behavior. Few intelligence tests, however, are ideal. Intelligence tests, whether for children or for adults, favor the tasks and situations experienced in schools. Intelligence tests deal with information, reading, vocabulary, arithmetic reasoning, computation, analogies, and the like. Insofar as the tests favor the kinds of things taught in school, they will tend to favor people who have had the greater extent of schooling. At best, an intelligence test measures a person's ability to deal with *aspects* of his environment at the time and up to the time of testing. It must be remembered, however, that an intelligence test measures ability to deal successfully with intellectual tasks. It is *not* an adequate measure of the individual's personality, adjustment, happiness, or vocational adaptation.

Higher scores on intelligence tests, however, can be interpreted as meaning better ability to deal with the intellectual environment. If higher scores can result from experiences in a more stimulating environment, or from the experiences in environments that favor the sort of things measured on intelligence tests, then society should seek to get the higher scores that do represent mastery of the intellectual environment. Intelligence tests show some of the consequences of the school environment. Some able individuals, however, are denied opportunity for maximum realization of their full potentialities when they cannot get all the education they desire or deserve or need.

FACTORS IN GOING TO SCHOOL

The extent of education is not solely related to intelligence. It depends upon economic circumstances, upon society's attitudes toward minorities, upon familial attitudes toward schools and learning, and upon society's ability to support education. These are but a few of the factors that are related to extent of schooling. Sibley[6] has shown not only that the most intelligent boys have a four-to-one advantage over the least intelligent in extent of education obtained, but also that the sons of men in the economically advantaged groups have a ten-to-one advantage over the sons of men in the lowest occupational groups. Davis[7] has shown that in the state of Kentucky about half of the best "college risks" did not enroll in college, while about a seventh of the poorest "college risks" did enroll. Among the best that did not go to college or the poorest that did go, the economic factor was the primary reason for *not* going among the ablest, or for going among the poorest. Karpinos and Sommers[8] definitely indicate that highest grade completed is dependent upon income level of the family, geographic region in which the person lives, and color. The facts revealed by the 1940 census[9] are evidence that economic circumstances and color are related to the extent of education. For adults twenty-five years of age and over, those resident in rural-farm areas average 7.7 years of completed schooling, whereas those resident in urban centers average 8.7 years. The average years of completed schooling for urban Negroes was 6.8 as contrasted with 4.1 for Negroes in rural-farm areas.

---

[6] Elbridge Sibley, "Some Demographic Clues to Stratification," *American Sociological Review*, 7:3, 322–30, June, 1942.

[7] Horace Leonard Davis, "The Utilization of Potential College Ability Found in June 1940 Graduates of Kentucky High Schools," *Bulletin of the Bureau of School Service, College of Education, University of Kentucky*, 15: No. 1, p. 101, September, 1942.

[8] Bernard D. Karpinos and Herbert J. Sommers, "Educational Attainment of Urban Youth in Various Income Classes," *Elementary School Journal*, 42:677–87, May, 1942; 766–74, June, 1942.

[9] Henry S. Shryock, Jr., "1940 Census Data on Number of Years of School Completed," *The Milbank Memorial Fund Quarterly*, 20:367–87, October, 1942.

The variation in the extent of schooling by region, state, color, and rural and urban residence suggests that many individuals in our society are denied through no fault of their own the opportunity for the fullest development of their potentialities. Since the evidence from this study indicates that limitations in fullest intellectual development may result from artificial restrictions in extent of education, society is losing tremendous amounts of its fundamental human resource. Educators, sociologists, and others must prevent the loss of society's human resource because of inadequate provision for its nurture.

SCHOOL ATTENDANCE vs. HIGHEST GRADE COMPLETED

It should be noted, however, that extent of education and amount of time spent in school are not synonymous. Keeping children in school to keep them out of the labor market or off the streets is not necessarily related to maximum development. In many ways it may even have undesirable results. A distinction must be made between *years of school attendance* and the *highest grade completed*. Thorndike[10] has shown that there is no relation between the number of years of school attendance and measured ability, but that there is a significant relationship between the highest grade completed and measured mental ability. He studied the forty boys of a group of more than seven hundred who had obtained the highest scores on a test of intelligence as contrasted with forty boys of the same group who had made the lowest scores on the intelligence test. He found that, on the average, the ablest forty were only four months older than the least able forty when they left school. The ablest forty, however, achieved high school graduation as contrasted with graduation from the eighth grade by the least able forty. Thorndike's study concerned itself with boys who were in school at a time when retardation was more prevalent than it is today. Some of the least able individuals in the group repeated grades twelve or more times, that is, they were "left back" or "held over" during their school careers an aggregate of six or more years. His study emphasizes that "going to school" and "getting something out of school" are not the same thing.

---

[10] Edward L. Thorndike, "The Distribution of Education," *The School Review*, 40:335–45, May, 1932.

The correlation obtained by Thorndike between number of years in attendance at school and measured intellectual ability of an adult is about zero. School attendance, which includes repetition of school grades for the least able and "skipping" of grades for the most able, is unrelated to an adult's measured mental ability. There is, however, a significant relation between the highest grade completed and an adult's mental ability. For the enlisted men in the United States Army during World War I, the correlation between highest grade completed and measured mental ability was between .60 and .70. The correlation suggests not only selective screening by schools, but also the increments to measured mental ability accruing as a result of education. Insofar as the 1921–22 tests of abstract intelligence represent a fair measure of ability under similar circumstances, the data indicate that increments in adult intelligence test score are related not only to original level of intelligence, but also to the extent of education that could be obtained. It should be obvious, therefore, that an intelligence test score of an adult cannot be interpreted as direct evidence of original inherent ability. On the contrary, the status of an adult on an intelligence test represents the interaction between his basic ability and the circumstances of his environment.

SUMMARY

Schooling makes a difference in a person's score on the intelligence tests that he takes as a mature adult. These differences are important not only for the individual but for society. Society must recognize that the restriction of educational opportunities because of race, color, and economic circumstance may mean the attenuation of its chief human resource—the functioning intelligence of its citizenry.

13

Deprivation, Development, and Diffusion

SUSAN W. GRAY AND RUPERT A. KLAUS

As American society has become increasingly concerned with poverty and the "culturally deprived," the consistent finding that the amount of schooling obtained is related to the intellectual level attained has led to a concern about the sizable proportion of the nation's school children who start out with below-average intelligence and who do not seem to profit very much from their school experience. As early as the 1930's, a controversial body of research results, the so-called Iowa studies, seemed to indicate that it is early *experience that has the most marked effect on intellectual development, and that preschool education of three-year-olds and four-year-olds can stimulate such development.*

With the general progress of knowledge about the effects of early experience on development, it was possible during the 1960's to set up research investigations that would be more definitive than the Iowa studies of thirty years before. One of these was undertaken at George Peabody College. The paper reprinted below is a report on the procedures and preliminary results by Susan W. Gray and Rupert A. Klaus entitled "Deprivation, Development, and Diffusion." It is an abridged version of Dr. Gray's presidential address to the Division of School Psychologists, Sept. 4, 1966, and is reprinted with the permission of the authors.

This afternoon I should like to talk with you about a research study just entering its sixth year, one that is concerned with the experimental testing of an intervention program designed to improve the educability of young deprived children. This study, the Early Training Project, to give it its short name, grew originally out of the concerns of the school administration in the city in which the Peabody program in school psychology maintains a field and practicum center for its students.

One problem, of several mentioned by the superintendent of schools, as the staff of the Peabody school psychology program talked with him about research possibilities, was that of the progressive retardation in achievement shown by the children in one of the city schools, a school which was all Negro, and in which as many as 75 per cent of the children came from deprived backgrounds. After some preliminary data gathering by students in the school psychology program, it became clear not only that such progressive retardation did exist, but also that the majority of the children were at a substantial disadvantage in the beginning, as shown on reading readiness tests and on first grade achievement.

And so, largely for practicum experience for students in planning research, in carrying it out, evaluating the results, and in gaining actual experience with young children, we designed a special ten-weeks intervention program to conduct with a group of deprived children for ten weeks prior to entrance into the first grade.

The results of this study intrigued us enough that we decided to plan a more extended one. And so, Rupert Klaus and I began in 1961 to plan a new intervention project with which we would begin much earlier with children, provide a more massive experience for them, and involve the mothers to a greater extent than we had done in the pilot study.

This study, The Early Training Project, from the standpoint of its design, the nature of the intervention program, and certain of its interim results, has been described in previous papers of this association and elsewhere. I should, therefore, like to spend my major time this afternoon in describing the long term results of the study—at least through the past summer. Since many of you, however, are probably not acquainted with the study, I hope others will forgive me if I take fifteen or so minutes to review the design of the study, the general rationale on which the intervention program was based, and the actual operation of the program.

Our general strategy grows out of the fact that at the time we began, research in the field had not yet demonstrated whether it is possible to offset substantially the progressive retardation commonly observed in deprived children in their years of schooling. It is true that several studies had demonstrated a 5 to 15 points IQ gain in young children with special interventions no longer than three months. These gains, however, tended to be temporary, and are probably explicable in part by the changes in motivation from pre-test to post-test associated with experiences in which children learn to respond to a range of adults, particularly verbally, and to become somewhat task-oriented.

Our effort was to design an "experimental package of best bets" for intervention. This package was based upon available research on social class, cognitive development, and motivation. From such research we attempted to derive those characteristics which appeared to be related to the differences in school performance of middle class and culturally deprived children and which also appeared to be subject to possible manipulation. Since we were concerned with school applications, as well, we attempted to do this within a framework it would be possible to duplicate, on a wide scale, should the project prove to be successful.

Our general aim was to make our intervention program *developmental* rather than *remedial*. Thus, we planned to begin at an early age. We set up four groups of children. Three of these groups were randomized from a group of 65 deprived children born in 1958 in a small southern city. One of these groups had three summers of a specially planned preschool, plus weekly visits during the remainder of the year by a specially trained home visitor whose task it was to involve the mother and child in activities similar to those of the summer experience and to give the mother some help in learning the instrumental steps involved in enabling her child to realize her aspirations for him, educationally and vocationally. A second group had a similar intervention schedule, except that it began one year later and thus involved only two summer schools. Both of these groups had the services of the home visitor during the first grade. The third group served as the local control group. A fourth group, in a town similar in economic structure, but 60 miles distant, served as an additional control group. This last group was added to the design since we anticipated the possibility of diffusion into he control group in the small city in which we worked; we oped by the additional group to arrive at some estimate of such iffusion effects. .

THE NATURE OF THE PROGRAM

The preschool summer experiences for each group were arranged so that each group of approximately 20 had a head teacher, plus three or four assistant teachers, the assistants being largely trainees in our school psychology program or undergraduates especially recruited for the program. . . . The high ratio of adults to children was of particular importance for the form of intervention we planned. The staff was about equally balanced as to race and as to sex. We made especial efforts to have men teachers in the group to serve as appropriate role models for the little boys, particularly those from father-absent homes. . . .

One of the two main classes of variables with which we were concerned in our summer program was that of attitudes related to achievement. Our paramount concern here was with achievement motivation and such related characteristics as the ability to persist and to delay gratification. During the first summer we found that gross motor activities appealed greatly to the youngsters and were one of the better ways to arouse their motivation to achieve or excel. . . .

Block building also served as a way of encouraging children to better their previous performances or those of others. Attention paid to the objects they were able to produce was something new to these youngsters. We mounted their pictures on the bulletin board, admired their little clay bowls, and their magazine cutouts.

The teachers served as reinforcing agents for the youngsters. Close physical contact, with the undivided attention of an adult, was one of the most effective reinforcements we found for the children. Father-figures were especially important for the little boys from father-absent homes, although the little girls as well enjoyed male attention. The adults thus could set the pattern of approved behavior for the youngsters and themselves serve as achieving role models for the children.

Since we were working with Negro children, we were particularly concerned with the problems of self-esteem which the Negro child experiences in the WASP world he encounters in the mass media, in school, and in the world at large. . . .

In terms of general cognitive development, we were first concerned that the children acquire some of the more elementary ways of classifying their world, of seeing similarities, and differences. . . .

Since the youngsters seemed somewhat retarded in perceptual discrimination, we worked with many types of material that would help increase their ability to see likeness and differences, to see relations of parts. . . .

Language was, of course, of major concern to us, and we used many ways of trying to develop it. Telephones were a natural for such efforts and greatly enjoyed by the children. . . . One inch cubes were put to use for learning position words and the like. . . . The old nursery rhymes were often used, and enacted to the children's delight. . . . We used songs, rhymes, and dramatization, not only to help develop language but also to promote a liking for school-type activities, most of which were foreign to the children's home situations. Most of all we read to the children, and presently had brought them to the place where looking at books became an absorbing activity.

We have been a little dubious about the term "enrichment" as applied to a project such as ours, since the term suggests throwing a lot of new experiences at children, without appropriate concern for whether the children have developed the ability to process the new information that comes their way. . . . We tried to insure that new experiences . . . should be paced and sequenced in such a way that the child could make use of the new information provided.

Some of these new experiences were as simple as using an easel and tempera paint, or in looking at one's hands through a tripod microscope. Even trying to blow up a balloon was a challenge to some of these youngsters. Rhythm instruments were highly attractive to the youngsters and useful for teaching such new meanings as together, faster, louder, softer, and the like. We took them to farms, where they had a chance to explore chicken houses, to see the baby calf, and to feed the pigs. A young man brought to school his mare and two-week-old foal that was bottle fed. Those who had the courage were allowed to take a ride on the mare. We visited an air base, a fascinating and somewhat overwhelming experience for these youngsters. I fear this was one time they got more information than they could process; they came home as if they had been on a lost weekend. We took them to the children's museum, which was a little beyond them, even during the third summer. The dioramas and the live animals, however, bore enough resemblance to the animals they had seen on the farm and in their picture books to intrigue them.

We took them to the library, somewhat with fear and trembling as to their behavior. The trip was a success, however, and was followed up by the home visitor. Prior to the visit, only two of the mothers had ever been to the library. After the trip, and particularly with the efforts of the home visitor, we were able to get all of the mothers except two to visit the library and to take out library cards. . . .

Over the five years that we have been working with the Early Training Project, we have managed to collect a whole mountain range of data—and the end is not in sight; we find the progress of our children and their families through school, in these times of rapid educational and social change, much too fascinating to let the study drop.

Before going into the analysis it might be helpful to describe briefly some of the observable differences in the four groups that do not relate to intervention but rather to the vicissitudes of randomization. . . . [Table 1] gives the breakdown of the children remaining in the four groups as they looked at the initiation of the study. Virtually all of the children are still with us. On none of the indices that you see were the groups statistically different, although the IQ difference between T_1 and T_2 was a squeaker. When we have gone back and examined the home situations, T_2 seems a favored group as compared to T_1. There are twice as many father-absent homes in T_1 as in T_2. Precise estimates of income are difficult to come by in these families.

TABLE 1

Status in May 1962 of Four Early Training Project Groups Used in 1966 Analyses.

	T_1 (Exper.)	T_2 (Exper.)	T_3 (Local Control)	T_4 (Distal Control)
CA (mo.)	45	46	47	45
Binet MA (mo.)	39	42	41	40
Binet IQ	86	91	87	88
PPVT MA	30	30	28	32
PPVT (raw score)	19	20	18	22

The best comparison we could make in 1964 indicated that the income, while low in both groups, was approximately double in T_2 what it was in T_1. This point would not be worth laboring, except that, with T_2 being a more favored group, it is difficult to make comparisons of the effect of the longer intervention with T_1 as compared to the shorter one with T_2. The only time that significant differences have consistently appeared between these two groups was prior to intervention. Because of the differences in the two groups, we believe that we can say very little, if anything, about the relative effectiveness of the two lengths of treatment which we employed. The other three groups appeared more similar on family indices. If we had to rank them in terms of general favorableness of the family situation, we would rank T_2 first, T_4 second, T_3 third, and T_1 fourth.

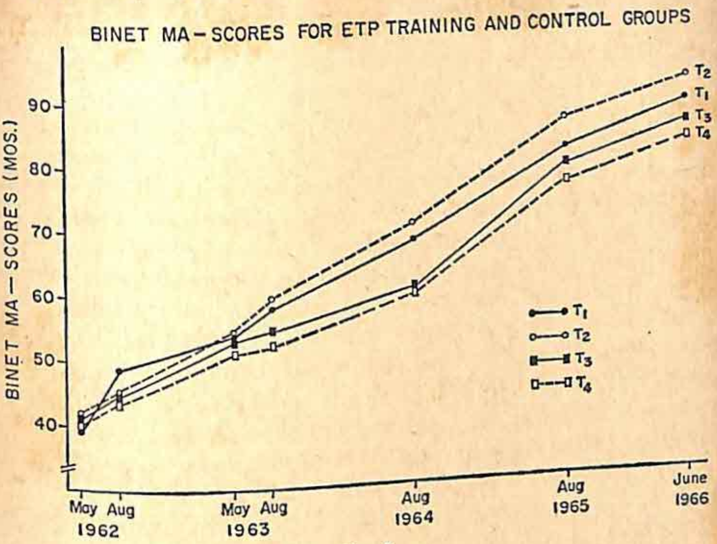
Probably a word should also be added about the differences in the two communities, since this becomes important in comparing local and distal control groups. The general economic situation in our major city is a little better (the population at least is increasing more rapidly), and the school system is slightly superior, as well as we can judge. On the other hand, the general economic situation of the Negro is better in our distal city, and has been so for many years, because of the presence of certain major industries that have hired Negroes in well-paying positions. We believe now, as we did at the time we chose the distal city, that it is the best match we could find. We have been troubled, however, by certain differences in the school system, and for this reason have tried where possible to make comparisons that would take this into account.

From 1962 through 1964, we have pre- and post-tested all four groups in the months before and after the special summer school. Routinely we have used the Binet and the Peabody Picture Vocabulary Test (PPVT). Starting in 1964, just prior to school entrance, we added the Illinois Test of Psycholinguistic Abilities (ITPA) and the WISC. These last two have been given once yearly, and the WISC has been alternated with the Binet. I should like first to go into our analyses of these data.

Our method of analysis has been to use Lindquist mixed designs, Type I, III, or VI, whichever was appropriate to our data. Where significant Fs occur, as they have in our data, we have used orthogonal comparisons to locate the sources of differences in groups. We have thus compared T_1 and T_2 combined with T_3 and T_4 combined, and then made additional orthogonal comparisons of T_1 and T_2, and of T_3 and T_4. These seem to us the most meaningful group comparisons for assessing the effects of our intervention program and of diffusion effects.

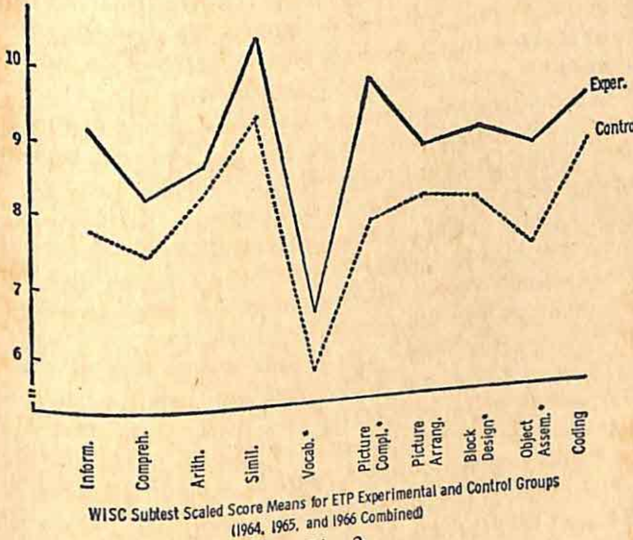
Now for the Binet. Figure 1 gives MA scores from May of 1962 to June of 1966, a period of 49 months. At all points since intervention began with T_2, we have found significant Fs for the comparisons of the experimental and controls. The general curves are ones of increasing divergence up to school entrance and then, in the school years, some closing of the gap. As you can see from the graph, the greatest difference in groups was just prior to school entrance. At this time T_1 had shown an IQ gain of 9 points and T_2 of 5 points, while the local and distal control groups had lost 4 and 7 points respectively. You can also see that the curves maintained approximately the same slope for the experimental groups during the first grade (August 1964 to August 1965) while both of the control groups showed some acceleration. Also apparent is the slight deceleration straight across the groups in the past year. None of the four groups in this last year made as many months of gain in mental age as might be expected with "normal" progress. This is a finding to which I wish to return later.

Now for the WISC (Figure 2). Since I thought you might be



BINET MA-SCORES FOR ETP TRAINING AND CONTROL GROUPS

Fig. 1.



WISC Subtest Scaled Score Means for ETP Experimental and Control Groups
(1964, 1965, and 1966 Combined)

Fig. 2.

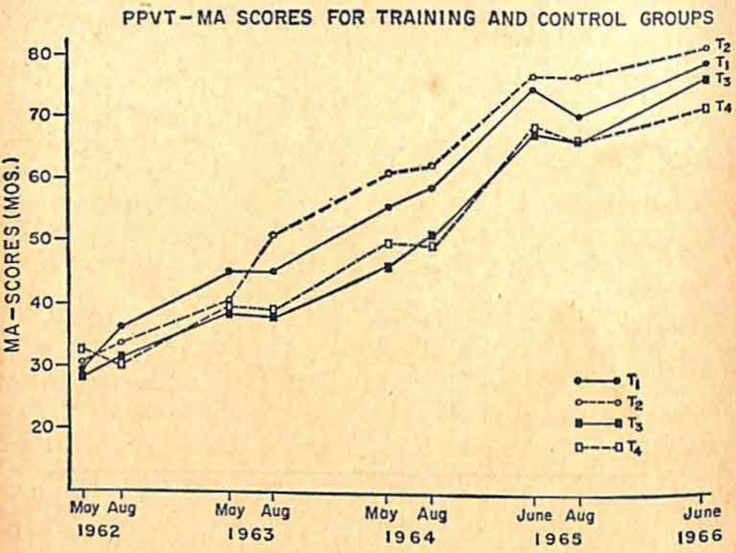


Fig. 3.

interested in the sub-tests, we have combined the three administrations. In general, the profiles are similar at all ages. Here you can see the striking deficiency in vocabulary, and the slight superiority shown on similarities. Experimentals are superior to controls on six sub-tests: information, arithmetic, vocabulary, picture completion, block design, and object assembly.

On the Peabody Picture Vocabulary Test (Figure 3), you see a graph somewhat similar to that for the Binet, in that you have a divergence of experimental and controls once intervention has begun, with considerable closing of the gap in the second year of schooling. In 1966, the experimentals are still superior to the controls, and the local control group is superior to the distal one.

[Figure 4 gives] results on the ITPA. . . . The three testings have been combined to give some notion of the general profiles. Here you can see that the differences are not great, and that the profiles of the experimental and control groups are almost parallel. The most conspicuous score is probably that on Auditory Vocal Automatic. Those of you who are familiar with the test

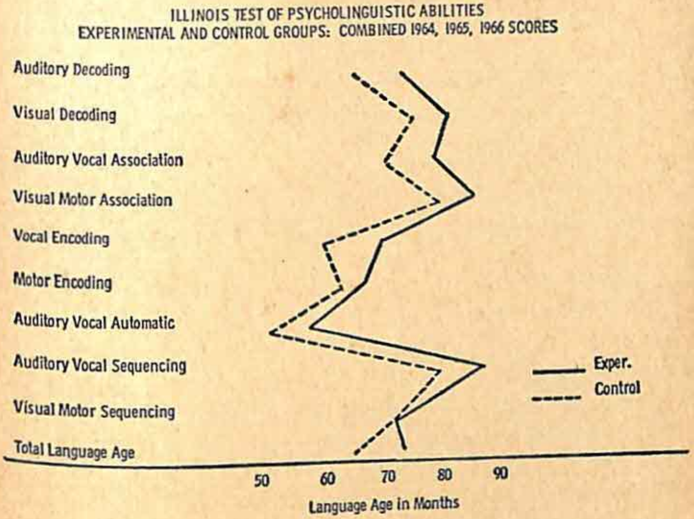

ILLINOIS TEST OF PSYCHOLINGUISTIC ABILITIES
EXPERIMENTAL AND CONTROL GROUPS: COMBINED 1964, 1965, 1966 SCORES

Fig. 4.

will recall that this sub-test demands the sounding of final consonants. The sub-culture from which these children come, lower-class southern Negro, is generally deficient in pronouncing final consonants; these children presumably are continuing the pattern acquired in their homes. In the Type VI Analysis of Variance which was performed, there was a significant group effect, a significant sub-test by administration, and a significant sub-test by group interaction. Sub-analyses indicated that, insofar as the effects of intervention go, differences are largely attributable to the performance of experimental and control groups on the 1964 administration; they do not differ appreciably thereafter.

So much then for the summary of our tests which have been given over time. From them I believe the picture is relatively clear that the maximum effect of intervention was obtained just prior to school entrance, exactly what we would expect. Although differences in experimentals and controls are still significant on the Binet, the PPVT and the WISC, the differences are of somewhat smaller magnitude.

But these tests were used largely as bench marks. In the Early

Training Project our concern has been chiefly with what happens in terms of school performance. Before trying to interpret these bench mark data, then, I should like to give you a picture of what has happened in the first two years of public schooling.

Like many others, I am somewhat skeptical of the adequacy of achievement tests in the first grade, and even in the second, as predictors of general school competence in the later years. Still, along with the measures already described, they are about the best we have available at present to answer the question of improvement of educability. . . .

Table 2 refers to first year performance on the Metropolitan Achievement Test . . . First, if you will, look at the bottom section of the table where the four treatment groups are compared, plus another group, which is included because it represents a comparison with children who presumably had little contact with the experimental and control children. You can see from these scores that typically the experimental children score a little higher than the local controls, and the local controls a little higher than the non-ETP group, and at about the same point as the distal control group. None of the differences in the two experimental and the two control groups are significant. Some of the differences, especially between the two control groups, are sizeable, but the large variance meant that differences are not significant, except in the case of word knowledge, where T_3 is superior to T_4.

Although the differences are not significant, their general patterning prompted us to inquire a little more into first grade performance in the schools, with particular concern for the possibility of diffusion from the experimental to the local control group. I might add that this general patterning showed up on practically every measure we took upon the children during the first year. A Kendall Coefficient of Concordance for thirteen measures upon the four groups was .86 ($p < .001$).

With the hunch that the introduction of 43 experimental children into the first grade of 91 children might have an effect upon over-all first grade performance, we next looked at performance scores on the Metropolitan Achievement Test, which were available at the two schools. These data you can see in the top two sections of the table. In the local school (where the 1962

TABLE 2

Early Training Project First Year Scores. Annual Median Grade Equivalent Scores on Metropolitan Achievement Test, Primary Battery; First Grade; (norm = 1.8).

	N	Word Knowledge	Word Discrim.	Reading	Arith.
Local School					
1961	61	1.69	1.77	1.72	1.64
1963	88	1.65	1.59	1.69	1.42
1964	89	1.63	1.58	1.71	1.57
1965	91	1.74	1.77	1.87	1.43
Distal School					
1962	47	1.60	1.64	1.61	1.47
1963	82	1.58	1.49	1.56	1.45
1964	81	1.47	1.42	1.54	1.41
1965	111	1.42	1.39	1.52	1.43
T_1 & T_2	43	1.71	1.73	1.89	1.61
T_3	18	1.65	1.75	1.75	1.38
Non-ETP deprived (local)	9	1.48	1.50	1.51	1.23
T_4	23	1.33	1.30	1.54	1.28

data were not available in usable form) scores appear to show no patterning until one reaches 1965, the year in which the experimental children were in first grade. In this case there does seem to be a general improvement, except in arithmetic. This is most conspicuous in reading, which is probably the best predictor of these four scores for future achievement.

In the distal school, another pattern emerges, which is one of a slight but consistent decline over the past four years. There is some informal evidence that this may be related to an increase in the number of children from relatively deprived backgrounds who have entered the school, but this is mostly guess work.

One thing that troubles us is the generally inferior performance of the distal school. As a partial check on the differences in schools, we went back to look at the performance of the older siblings of the children in the two experimental and the two control groups, on the assumption that differences in the general adequacy of the schools should appear in the older siblings' test scores, and that these would be relatively independent of any treatment effect of our intervention. This we did, and calculated gain scores for these children, from a first to second testing on the Metropolitan. In this comparison there are no significant differences in the amount of gain shown by the older siblings of T_1 and the local and distal control groups. The older siblings of T_2, however, made significantly more gain, part of the picture of the greater adequacy of the T_2 families already mentioned. Since T_1 and T_3 showed the same patterning as T_4, we are inclined to think that we are not dealing simply with a comparison of two school systems but may be dealing instead with a phenomenon of spread of effect, which we have termed horizontal diffusion. I should like to return to this in a moment, after we have looked at achievement test scores for the second year.

First for the results on the Stanford Achievement Test (Table 3). Three things, I believe, emerge from this table. First, and regrettably, none of the groups is performing up to grade expectation, since at this point they should have been at approximately 2.8 in mean grade equivalent. Secondly, the experimental children are superior to the control children in four of the six tests, as can be seen at the bottom of the slide. Most of the variance here, however, was contributed by the distal control group. On these same four tests, the local control group is superior to the distal control group.

TABLE 3

Stanford Achievement Test Scores Mean Grade Equivalent, May 1966.

	Word Reading	Para. Meaning	Spelling	Word Study Skills	Arithmetic	Vocabulary
T_1 (Exper.)	2.2	2.4	2.2	2.0	1.9	1.9
T_2 (Exper.)	2.2	2.4	1.9	2.2	2.2	1.9
T_3 (Local Control)	2.1	2.2	2.2	2.3	2.1	1.9
T_4 (Distal Control)	1.7	1.7	2.0	1.6	1.7	1.7
	$(T_1 + T_2) >$ $T_3 + T_4$ $T_1 - T_2$ $T_3 > T_4$	$(T_1 + T_2) >$ $T_3 + T_4$ $T_1 - T_2$ $T_3 > T_4$	NS	$(T_1 + T_2) >$ $T_3 + T_4$ $T_1 - T_2$ $T_3 > T_4$	$(T_1 + T_2) >$ $T_3 + T_4$ $T_1 - T_2$ $T_3 > T_4$	NS

TABLE 4

Metropolitan Achievement Test Scores, 1966. Mean Grade Equivalents and Gains from 1965 Administration.

		Word Knowl.	Word Disc.	Reading	Arithm.
T_1 (Exper.)	Mean Gain from '65	2.3 .6	2.6 .9	2.5 .7	2.4 .8
T_2 (Exper.)	Mean Gain from '65	2.5 .8	2.7 1.0	2.8 .8	2.6 .9
T_3 (Local Control)	Mean Gain from '65	2.3 .6	2.6 .9	2.6 .8	2.5 1.1
T_4 (Distal Control)	Mean Gain from '65	2.0 .6	2.2 .9	2.1 .6	2.0 .8
Significant Differences in Means		$(T_1 + T_2) >$ $(T_3 + T_4)$ T_1-T_2 T_3-T_4	NS	$(T_1 + T_2) >$ $(T_3 + T_4)$ T_1-T_2 T_3-T_4	$T_3 > T_4$

Now for the Metropolitan Achievement Test results (Table 4). . . . [This table] shows much the same picture as was shown with the Stanford Achievement Tests insofar as the relative status of the several groups is concerned. Again, most of the variance is contributed by the fourth group, although it is only on arithmetic that it is inferior to the local control group. An additional feature of this table is that it shows the mean gain since the time that the children were tested a year earlier. The gain scores suggest a slightly less than average progress on the part of all these groups. In T_2 and T_3 it is close to normal, approximately .9 of a year, while it is slightly less in T_1 and T_4.

How do we make sense out of these findings over four years since we began intervention, and over two years of public school?

We can begin with the perpetual cry of the psychologist that we need more data before we are able to draw adequate conclusions. Indeed we do, particularly since first and second grade achievement tests are not the best predictors. But let me at least be rash enough to try to interpret our data and at the same time to give you a little more information that may shed some light on these findings.

First, we might try to explore possibilities as to why the local control group in general performs more adequately than the distal control group. Differences in school systems are a distinct possibility, although you may recall that we did not find differences in gain scores for older siblings between the two schools, with the one exception of the T_2 group. The other possibility is that of diffusion into the local control group. The data you have seen, particularly the jump in first grade performance in 1965 in the local school, suggest this possibility. This we might term a product of diffusion. We have also been interested in the process of diffusion within the community. Our home visitor has supplied us with some evidence, mostly of an anecdotal character. My favorite is about the mother of twin boys in the local control group. Although she had only a passing acquaintance with the woman next door, when that woman's child was selected for the first experimental group, she became a daily visitor. Whatever things the experimental child brought home from school, or whatever the home visitor left with her, such as books and drawing materials, the twins' mother bought for them. They are now very competent little boys. In a little more data-oriented way, we have, through our home visitor's contacts, been able to establish fairly well certain patterns of kinship among the three local groups. . . . [Figure 5] shows that all groups are related, although T_1 and T_2 are more closely so than is either group with T_3. After two years in the homes of the children, the home visitor was also in a position to estimate with fair accuracy the number of sustained contacts between experimental and control families. This is shown in [Figure 6]. . . . Here you can see that there is only one child in the local control group from a family that has no direct or indirect contact with an experimental family. In six cases, a child has at least two direct contacts.

We have also been interested in the process and product of diffusion within the family group, which we have called vertical diffusion to distinguish it from the diffusion across groups, which we have termed horizontal. We have already analyzed data on 57 younger siblings of the Early Training children; the results of this analysis are shown on the next table (Table 5). These results do not admit of an easy interpretation. One might expect that, if diffusion is taking place, the younger siblings should be superior to their matches in the experimental groups. Instead,

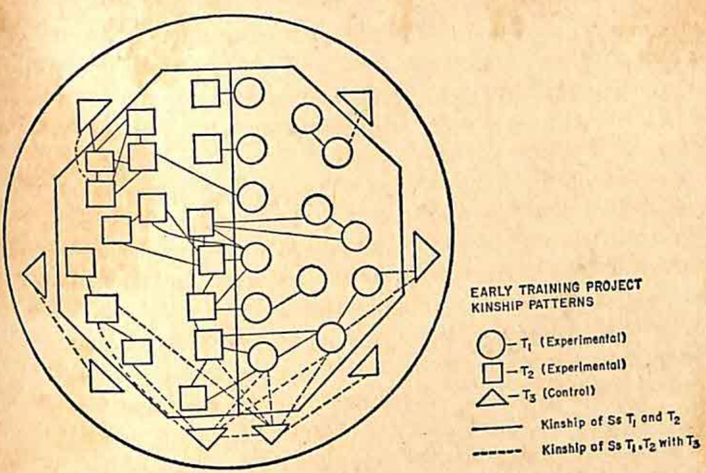

EARLY TRAINING PROJECT
KINSHIP PATTERNS
○ – T_1 (Experimental)
□ – T_2 (Experimental)
△ – T_3 (Control)
Kinship of Ss T_1 and T_2
Kinship of Ss T_1, T_2 with T_3

Fig. 5.

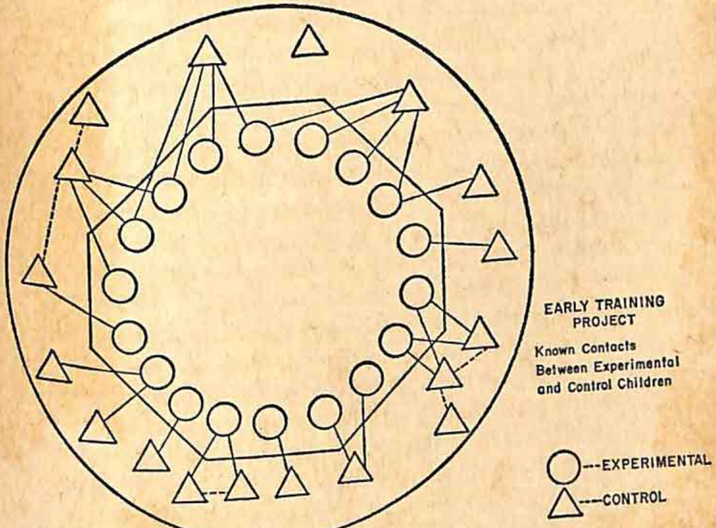

EARLY TRAINING
PROJECT
Known Contacts
Between Experimental
and Control Children
○---EXPERIMENTAL
△---CONTROL

Fig. 6.

control group in general performs more adequately than the distal control group. Differences in school systems are a distinct possibility, although you may recall that we did not find differences in gain scores for older siblings between the two schools, with the one exception of the T_2 group. The other possibility is that of diffusion into the local control group. The data you have seen, particularly the jump in first grade performance in 1965 in the local school, suggest this possibility. This we might term a product of diffusion. We have also been interested in the process of diffusion within the community. Our home visitor has supplied us with some evidence, mostly of an anecdotal character. My favorite is about the mother of twin boys in the local control group. Although she had only a passing acquaintance with the woman next door, when that woman's child was selected for the first experimental group, she became a daily visitor. Whatever things the experimental child brought home from school, or whatever the home visitor left with her, such as books and drawing materials, the twins' mother bought for them. They are now very competent little boys. In a little more data-oriented way, we have, through our home visitor's contacts, been able to establish fairly well certain patterns of kinship among the three local groups. . . . [Figure 5] shows that all groups are related, although T_1 and T_2 are more closely so than is either group with T_3. After two years in the homes of the children, the home visitor was also in a position to estimate with fair accuracy the number of sustained contacts between experimental and control families. This is shown in [Figure 6]. . . . Here you can see that there is only one child in the local control group from a family that has no direct or indirect contact with an experimental family. In six cases, a child has at least two direct contacts.

We have also been interested in the process and product of diffusion within the family group, which we have called vertical diffusion to distinguish it from the diffusion across groups, which we have termed horizontal. We have already analyzed data on 57 younger siblings of the Early Training children; the results of this analysis are shown on the next table (Table 5). These results do not admit of an easy interpretation. One might expect that, if diffusion is taking place, the younger siblings should be superior to their matches in the experimental groups. Instead,

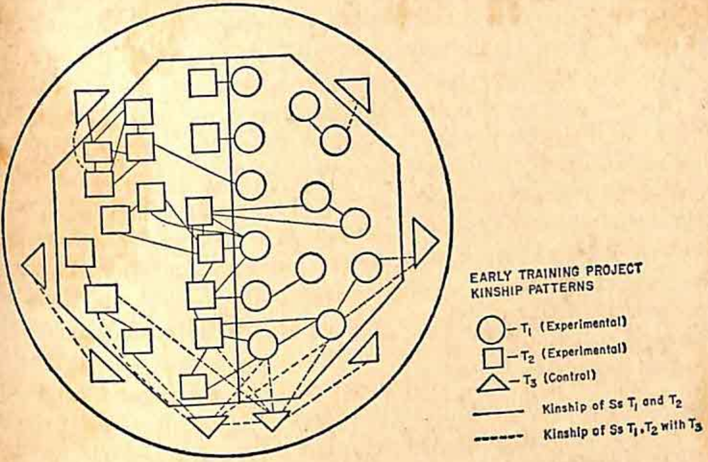

EARLY TRAINING PROJECT
KINSHIP PATTERNS
○ — T_1 (Experimental)
□ — T_2 (Experimental)
△ — T_3 (Control)
Kinship of Ss T_1 and T_2
Kinship of Ss T_1, T_2 with T_3

Fig. 5.

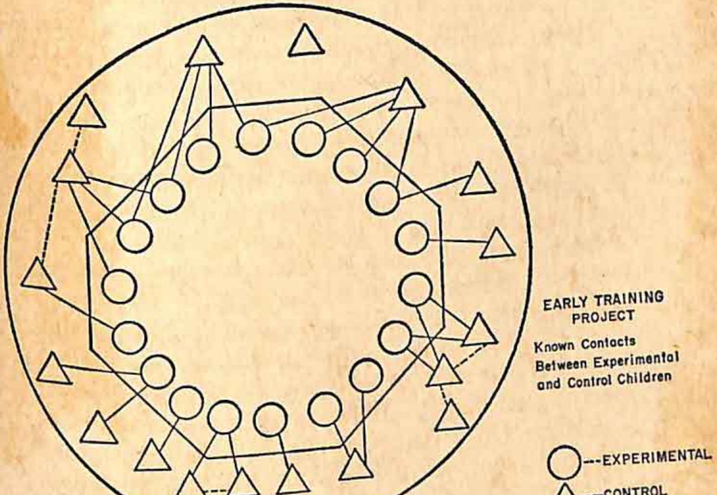

EARLY TRAINING PROJECT
Known Contacts Between Experimental and Control Children
○ ---EXPERIMENTAL
△ ---CONTROL

Fig. 6.

TABLE 5

Binet IQs of Early Training Project. Children with Younger Siblings of Testable Age, 1964.

	T_1 (Exper.)			T_2 (Exper.)			T_3 (Local Control)			T_4 (Distal Control)		
	N	CA (Mos)	IQ	N	CA (Mos)	IQ	N	CA (Mos)	IQ	N	CA (Mos)	IQ
Total Group	22	45	86	21	48	91	19	47	87	26	46	87
Project Children with Younger Siblings (Initial Testing)	12	47	82	16	47	89	7	50	84	12	48	88
Younger Siblings (Initial Testing)	13	54	82	21	53	83	9	54	71	14	62	74

scores are the same in the first group (which had the longest period of home visitor contacts) while a 5 point discrepancy is shown in the second group (with the shorter intervention period). Both of the control groups, however, show discrepancies of at least 13 points. These differences are significant. Our guess, which we have not been able to test adequately as yet, is that the younger siblings' scores, across the board, are slightly depressed by three possible factors: (1) they were older at the time they were tested, as you can see from the table; (2) there may have been a consistent examiner bias since the same person tested all four groups; and lastly (3) we have a notion, on which it is difficult to find or gather hard data, that ordinal position may be a variable in test performance in deprived families. Each additional child places a heavy drain on the family's resources of money, time, energy, and affect. We have just finished testing 77 yet younger siblings of the project children. We hope that our analysis of their scores will shed a little more light on this problem.

Next, and finally, we might look at what I am inclined to call dispersion, to follow the three D's of my title. Does the effect of a preschool intervention program inevitably wear off over time? Some people claim this and offer data to substantiate it. Other persons, such as David Weikart, have some hard data suggesting an effect that lasts well into the school years. I suspect it depends on nature of the program. We believe that our data—so far—show two things. One is that a program of preschool intervention as extended as ours and as tailored to the problem can have lasting effects. Our results on the Binet, the PPVT, and the WISC substantiate this. There is some suggestion of it in the other data, although it is less clear cut. Secondly, we believe that our data have something to say about the relevance of the nature of the public school and the community in which it is imbedded for maintaining the gains made during an intervention program. The schools to which most of our children go are good schools in many respects. In others, they leave something to be desired. All of our distal children and all but nine of our local children are in schools in which during the past year all children and almost all teachers were Negro. They continue to live in the same segregated communities, which have not changed greatly over the five years we have been testing the children. We are inclined to think that two things may be important here. One is that the motivational and cognitive patterns which we have tried to instill in our experimental children have not been sufficiently capitalized upon by the schools, since these schools deal largely with children that were like ours before our intervention project began. The other is that in these schools there appears to be somewhat more concern for the outward appearance of the school and its pupils than there is for the learning of the children. This is characteristic of schools in deprived neighborhoods; ours are not unusual. It does mean, however, that achievement may suffer.

TABLE 6

Metropolitan Achievement Test Scores: Mean Gains from 1965 to 1966 of Two Groups of ETP Children.

	N	Word Knowl.	Word Disc.	Reading	Arithm.
ETP-Ss who transferred to a White School in 1965	9	.96	1.45	1.35	1.40
ETP-children who remained at Negro School matched on IQ to the first group	9	.72	.86	.89	.72
Differences in favor of transferred group		.24	.59	.46	.68

As one check on this possibility, we thought it would be interesting to compare the gain scores of the nine children who transferred to previously all white schools in the fall of 1965 with a group of nine children who did not transfer but who were matched on IQ, and, as best we could, on parental aspirations. These findings are shown in . . . Table 6. With the small number of cases, and the difficulty of providing an adequate match in terms of motivational variables, I do not put too much credence in our findings. Nevertheless it is interesting to see that all differences favor the transferring children, and, that except in word knowledge, differences are sizeable.

Not only the school, but also the home, must be considered. Certainly, if no massive changes are made in the home conditions of children, it would be surprising if the situation that created the original deficit did not continue to take its toll. Our home visitor did work with each mother in a weekly session. Once a week for only two or three years, however, is hardly enough to bring the mother up to providing a home situation analogous to that of a more favored middle class child, who is the one who makes the "normal" progress on aptitude and achievement tests.

We began the Early Training Project with the question as to whether it is possible to offset the progressive retardation usually observed in deprived children as they go through the years of elementary school. We have not answered the question, *yes* or *no*. It probably can only be answered, "It depends." It depends first on the intervention program, but the intervention program cannot carry the whole burden. We suspect that it is possible to offset progressive retardation by a carefully designed preschool program—ours at least has lasted through the second year of school—but that it is folly to hope that such a program can continue to offset inadequate schools in later years. Our findings on horizontal and vertical diffusion make us hopeful that intervention programs can have long lasting effects that go beyond the immediate children with whom one may be working. Such programs, however, cannot be expected to carry the whole burden of providing adequate schooling for children from deprived circumstances; they can only provide a basis for future progress in schools and homes that can build upon this early intervention.

14

The Inheritance of Mental Ability

Cyril Burt

Research indicating that the process of mental development is modifiable and that a person's intellectual capacity depends to some extent on the kind of intellectual "nourishment" available to him as he grows toward maturity does not prove that individual differences in intelligence result only *from environmental differences. In fact, there is a large and growing body of evidence that heredity plays a major part in such differences. The paper reprinted here is a wise and comprehensive survey of much of this evidence by an English psychologist, Cyril Burt, whose own research over a long and productive lifetime added much to what we know about intelligence and how it develops. It is a slightly abridged version of "The Inheritance of Mental Ability,"* American Psychologist, *1958,* 13, *1–15. It is reprinted with the permission of the author and the American Psychological Association. (At the suggestion of the author, Table 1 has been replaced by another table based on a larger number of subjects, resulting from additional work since 1958.)*

THE PROBLEM OF INDIVIDUAL DIFFERENCES

In Britain as in America, the earliest applications of the new techniques of mental testing were concerned chiefly with the lower end of the intellectual scale. They were readily accepted as practical aids to the doctor in diagnosing the mentally deficient, and to the teacher in discriminating the temporarily retarded from the irremediably dull. It was Bingham's firm belief that: "in the long run it would be even more profitable to discover and aid the bright and the supernormal than to ascertain and provide for the dull and defective."

In this country the traditional method of dealing with the problem was by means of what was popularly known as "the scholarship system." British teachers and parents were apt in those days to think of scholarships rather as a means for rewarding hard work than as a device for detecting and assisting impecunious talent. But considerable misgiving was felt over the erratic local distribution not only of scholarship winners but also of defectives. It was indeed widely affirmed that these variations in the numbers were "largely the result of the injustice of the ordinary methods of examination." Accordingly, in 1913 the London County Council took the unprecedented step of appointing an official psychologist. His chief duties were to act as referee for what were sometimes known as "problem children"—particularly in disputed cases of feeble-mindedness and of scholarship ability.

In the surveys that I carried out in this capacity I found that, in the poorer districts of the East End, there were three times as many certified defectives, eight times as many backward pupils, but only one-tenth the number of scholarship winners as were reported in the well-to-do areas of London. Who or what was to blame: the teachers, the homes, the traditional examination in arithmetic and English? Or the fact that the varying modicum of ability bestowed on each child was irretrievably fixed at birth? Environment or heredity, those seemed to be the alternative explanations; and each had its own enthusiastic advocates. As a rule the social reformers gave one answer and the psychologists the opposite.

During the nineteenth century the main champions of social reform were the philosophical radicals: Bentham, James and John Mill, and their various followers. Their motto was the maxim of Helvetius: *l'education peut tout;* or, as James Mill put it, "if education cannot do everything, there is hardly anything it cannot do." In psychology, their one basic law was the law of association; and the wide differences between one individual and another they attributed solely to the cumulative effects of association, through the operation of "habit, custom, training, and environmental opportunity or deprivation."

By the beginning of the twentieth century, however, most British psychologists had abandoned this simple creed. Impressed by the novel doctrines of the evolutionary school, they began to follow Darwin, Spencer, Huxley, and Galton in holding that the laws of heredity and individual variation applied to men as well as to animals, to mind as well as to body. "It is not so much," said Karl Pearson, "that the slums create the dullards, but rather that the duller stocks gravitate automatically to the slums." It followed that the only cure would be to start breeding citizens as we breed race horses, prize rabbits, or pedigree pups. Huxley's celebrated essay on the "Natural Inequality of Man" won over the few who still wavered. Most teachers were already converted by firsthand experience under the new scheme of universal education.

However, during the last few decades psychologists have been gradually discovering that the problem is far more complicated than either side had originally assumed. And, largely as a result, there has been yet another swing of the polemical pendulum. In part this is due to a reluctance to support what G. K. Chesterton once described as "the modernized dogma of predestination—a dogma calculated to paralyse all progress towards a welfare state"; in part it has been due to a fuller appreciation of the possibilities of experience and learning, which in turn has involved a reversion to something like a rejuvenated belief in the fundamental law of association.

To the British psychologist the experimental work on "conditioning" seemed to offer a more precise interpretation, on a more adequate basis, of the traditional doctrine of association as preached by Hartley and Mill; and the behaviourist school was widely welcomed as substituting a more profitable line of practical research for inconclusive speculations about heredity and innate endowment. As a result, we find contemporary educationists reminding each other of Bagley's well known dictum that, "except for a few cases of pathological deficiency, the factor of heredity plays a very small part in human life, as compared with the factor of environment" (*1*); while several of their psychological colleagues quote with approval the pronouncements of Watson: "there are inheritable differences in structure, but we no longer believe in inherited capacities." "Give me," he adds, "a dozen healthy infants, and my own world to bring them up in, and I'll guarantee to train any one of them to become any type of specialist I might select—doctor, lawyer, artist, merchant-chief, and even beggar-man or thief" (*38*, p. 104).

A glance through the current literature will show that, for the most part, the hypotheses which our younger psychologists desire to challenge are the two propositions which, at the beginning of the century, were accepted almost without question as justifying the organization of our system of education: first, the hypothesis of general ability, i.e., the view that "up to adolescence the chief differences distinguishing individual pupils are differences in 'general intelligence' "—a view which led the Chief Inspector of Schools to declare that, "with a few exceptions, each pupil may be taught in the same class for every subject, or, if mentally defective, sent to a special school on the ground that he is defective all round"; secondly, the hypothesis of mental inheritance, i.e., the view that the wide differences in the degree of general ability exhibited by various pupils were due mainly to genetic endowment.

The Hypothesis of General Ability

Of these two hypotheses the first was firmly rejected by the report of the "Norwood Committee"—a report which strongly influenced the reorganization of our national education as envisaged in the Education Act of 1944. The committee maintained that children differed far more in quality of ability than in amount; and recommended a tripartite classification of "secondary" schools, based on the doctrine that pupils fall into three main types: a literary or abstract type, to be educated at "grammar schools"; a mechanical or technical type, to be educated at "technical schools"; and a concrete or practical type to be educated at "modern schools." This conception was supported by several leading psychologists who argued that "the statistical analyses of Thurstone disclose no evidence for Spearman's factor of general ability, but only a number of 'primary abilities,' each more or less specialized" (*8* and refs.).

The scheme has not worked out in actual practice so successfully as was hoped; and the proposal itself at once stimulated a large number of new inquiries. Those who still defend the tripartite theory seem to overlook the vast and varied character of the evidence that has since become available on the issue thus raised. As a rule they content themselves with citing one particular line of reasoning—usually the statistical. If, however, we take a wider view, we shall find that there are three or four independent chains of converging arguments, none perhaps irrefragable by itself, but all of them tending to rehabilitate the older view. The evidence on which these inferences rest is drawn from many different fields: introspection, observation, biology and neurology, as well as statistical research.

In point of fact, the distinction between "general ability" and "special abilities" was due, not (as is so widely supposed) to Spearman but rather to Francis Galton, from whom Spearman, as he expressly states in his earlier papers, originally borrowed it. Spearman, however, went on to deny the existence of "special abilities" which Galton had supported, on the ground that they were "just relics or revivals of the obsolete theory of faculties" (*31*). It is this aspect of the Spearman doctrine that the modern educationist repudiates, and with it he is a little too ready to discard "general intelligence" as well.

However, even Galton has been partly anticipated by previous writers, who, in a vague and inchoate fashion, had inculcated a somewhat similar doctrine on the basis of speculative arguments, partly introspective, partly biological. Introspection, so the faculty school had claimed, revealed a sharp distinction between intellectual or cognitive qualities on the one hand and emotional or orectic qualities on the other. Spencer and his followers accepted this distinction, and supplemented it with biological arguments, which in those days were almost equally conjectural. During the evolution of the animal kingdom, and again during the development of each individual, the fundamental capacity which Spencer termed intelligence

> . . . becomes [so he tells us] progressively differentiated into a hierarchy of more specialized cognitive capacities—sensory, perceptual, associative, and ratiocinative, much as the trunk of a growing tree ramifies into boughs and twigs (*32*, Part iv).

Later writers were not so sure that the term cognition really furnished the best differentia: for them the real distinction was between directive or adaptive processes on the one hand and valuative or dynamic processes on the other; "intelligence," said Sully, "steers like a rudder; emotion and interest supply the steam." A similar reinterpretation has more recently been proposed by Piaget in his book on *The Psychology of Intelligence* (27, pp. 4ff.).

Far more convincing to my mind is the neurological evidence. The alternative doctrine of distinct and independent abilities drew much of its support from early experiments on cerebral localization. These, it was claimed,

> . . . appear to indicate that distinct functions—motor, sensory, perceptual, associative, linguistic, and the like—can be assigned to distinct and definite centres or areas within the brain.

And this conclusion was apparently corroborated by the maps of cell structure subsequently produced by histologists like Campbell, Brodmann, and von Economo. Their inferences, however, have been severely criticized by more recent workers, like Bok, Lorente de Nò, and, at University College, by D. A. Sholl (*30*). Few of the older map makers (he says) realized "the enormous amount of variability that exists between individual human brains." In their studies of comparable cortical areas from different specimens of *Ateles* Lashley and Clark have shown that the differences between different individuals are at least as large as the differences upon which the distinction of "architectonic areas" was based (*23*). Hence it now seems clear that the charts that used to figure in popular textbooks "greatly exaggerated the amount and the definiteness of the alleged localization, both as regards structure and as regards function."

After all, as Sherrington so frequently pointed out, in any one individual each anatomical tissue—skin, bone, hair, or muscle—tends to be of the same general character all over the body: minor local variations are often discernible; but in the same individual the variations are much slighter than those between one individual and another; and we should naturally expect the same to hold good of nerve tissue. Bolton's studies of the cerebral cortex, both in normal and defective persons, suggest that the quality of the nervous tissue in any given individual tends to be predominantly the same throughout: in low-grade defectives, for example, the nerve cells tend to be "visibly deficient in number, in branching, and in regularity of arrangement in every part of the cortex." The clinical work of Hughlings Jackson, the experimental researches of Sherrington, and the microscopical studies of the infant brain carried out more recently by Conel and de Crinis, seem in many ways to provide increasing confirmation for Spencer's theory of a "hierarchy of neural functions," developing stage by stage out of a simpler basic activity into higher and more specialized forms. Even if we cannot wholly accept Lashley's doctrine of the "mass action" of the brain, nevertheless the actual facts that he has reported are far more in keeping with a theory of general ability (and minor special abilities) than with a theory of major "special abilities" and no general factor. And Sholl himself concludes that "intelligence," like retentivity, is a "general attribute of all cortical tissue" (*30*, p. 111).

Each of the foregoing arguments—from introspection, from biology, and from neurology—is admittedly inconclusive. The most we could so far claim is that, when taken in combination, the evidence from each of these separate fields sets up a fairly strong a priori probability in favour of Galton's twofold theory of general abilities plus special. It is at this point, therefore, that the psychologist is tempted to invoke the aid of statistical analysis. In my view, the function of statistics, at any rate in psychological research, is simply to provide a more rigorous method of testing alternative hypotheses. How then do the various hypotheses that confront us stand up to this ordeal?

Each of the three main suggestions leads to distinctive corollaries which are open to statistical verification. First of all, if, as Thorndike held, there was no such thing as a general factor, but only a set of independent abilities, then the cross-correlation between the tests or assessments for those abilities should be zero, or at least nonsignificant. They are not: they are always both significant and positive. Secondly, if, as Spearman maintained, there were no such thing as special abilities but only a single general factor, the intercorrelations for every test should diminish proportionately in the same descending order (except, of course, for nonsignificant deviations due to sampling and the like): in technical language, the table of intercorrelations should constitute a matrix of rank one. It does not.

The third hypothesis assumes the presence of both general and special abilities. In the early article (*3*) to which I have already referred, I endeavoured to apply to the correlations between mental tests the procedure which Karl Pearson had already elaborated for use with measurements of bodily characteristics. In this way I sought to compare each of my various tables of observed coefficients with the best fitting set of theoretical values deduced from the hypothesis of a single "general ability" only; this hypothetical ability was redefined as the "highest common factor" entering into all the processes measured in any given table. The fit was moderately good, but by no means perfect. The differences between the observed and the theoretical values were accordingly tested for statistical significance; and the result was to disclose small but significant clusters of residual correlations, common to certain limited groups of tests only, thus plainly indicating the presence of certain supplementary abilities. This seemed definitely to clinch the double hypothesis put forward by Galton. Precisely the same conclusion was reached in many later analyses carried out both by myself and by other workers with larger numbers of pupils and a greater variety of tests (*5*, *6*, and refs.).

The factorization of correlations between persons and between successive occasions (with the same persons and tests) revealed, or at least very plainly confirmed, the constant presence of a random factor, which appeared to be due, not merely to sampling as ordinarily understood, but to something essentially characteristic of the working of the mind or brain. A somewhat similar conclusion has been independently reached by neurologists like Beurle, Cragg and Temperley, and Uttley and Sholl. Thus the type of "neuronal connectivity" suggested by the histological study of dendritic fibres strongly suggests "the semi-random arrangement of a machine working largely in accordance with the principles of probability" (*30* and refs.). This partly agrees with, though it is by no means identical with, the doctrine of "equipotentiality" put forward by Lashley; and thus further supports the value of factorial procedures as a means of investigating variations in mental ability.

However, what is far more important from the practical standpoint of educational guidance is this: in nearly every factorial study of cognitive ability, the general factor commonly accounts for quite 50% of the variance (rather more in the case of the young child, rather less with older age groups) while each of the minor factors accounts for only 10% or less. With increasing age, the "group factors," which represent special abilities, contribute increasing amounts to the total variance; and this fact lends further support to the view that, as a result of maturation, cognitive ability progressively differentiates and tends to become more and more specialized (10 and refs.).

Be that as it may, whatever be our views on the various biological or neurological issues, we may, I think, safely say this: for purposes of prediction—forecasting what this or that individual child is likely to do in school or in after-life—the general factor is by far the most important, though admittedly not our only, guide. For all practical purposes, almost every psychologist—even former opponents of the concept of general intelligence, like Thorndike, Brown, Thomson, and Thurstone—seems in the end to have come round to much the same conclusion, even though, for theoretical purposes, each tends to reword it in a modified terminology of his own. And thus today, save for one or two occasional attacks, current psychological criticism is not so much concerned with the problem of "general intelligence," but is directed rather against the second of our two initial propositions: the assumption, namely, that individual differences in intelligence are hereditary or innate.

The Genetic Component

Here three distinct questions seem to be involved: (a) what evidence is there for the *fact* of inheritance, (b) what precisely is the *mode* in which intelligence is inherited, and (c) what is the *relative importance* of the genetic factor as compared with the environmental?

The Fact. In controversies about the facts of mental heredity most critics have tended to assume that the two causal agencies commonly discussed—heredity and environment—are not merely antithetical but mutually exclusive. The environmentalists apparently suppose that, once they have shown that intelligence tests are affected by environment, it follows that all differences in intelligence are due to nothing but environment. Similarly the thorough-going hereditarian is apt to talk as though he believed that differences in intelligence were due to nothing but genetic constitution. This is the old familiar fallacy which I am tempted to label "nothing-buttery." In point of fact, with a few rare exceptions, like eye colour or serological differences in the blood, every observable characteristic that geneticists have studied has proved to be the product of the joint action of both heredity and environment. There are, in short, no such things as hereditary characters: there are only hereditary tendencies.

Now, where two contributory factors, such as heredity and environment, are likely to be involved, the obvious procedure will be to keep first one and then the other as constant as possible, and observe the results in either case.

1. *Uniform environment*. As psychological consultant to the London County Council, I had free access to orphanages and other residential institutions, and to the private files of case records giving the history of the various inmates. My co-workers and I were thus able to study large numbers of children who had been transferred thither during the earliest weeks of infancy, and had been brought up in an environment that was much the same for all. To our surprise we found that individual differences in intelligence, so far from being diminished, varied over an unusually wide range. In the majority of cases, they appeared to be correlated with differences in the intelligence of one or both of the parents. Some of the most striking instances were those of illegitimate children of high ability: often the father (so the case records showed) had been a casual acquantance, of a social and intellectual status well above that of the mother, and had taken no further interest in the child. In such cases it is out of the question to attribute the high intelligence of the child to the special cultural opportunities furnished by the home environment, since his only home has been the institution.

2. *Uniform heredity*. To secure cases in which the children's genetic endowment is the same, we may turn to assessments ob-

tained from monozygotic or "identical" twins. The mother is not infrequently unable or unwilling to bring up two children at the same time, and one twin is consequently sent to a relative or to a foster home. Owing to the strong popular prejudice against separating twins, she not unnaturally tries, as a rule, to keep these arrangements secret. But patient and tactful inquiries show that cases of twins brought up in different environments almost from birth are in fact much commoner than is usually believed. We have now collected over 30 such cases (*9*, *11*). I reproduce the more important correlations for the twins in Table 1 and have added for comparison corresponding coefficients obtained from other pairs, both related and unrelated. As regards intelligence the outstanding feature is the high correlation between the final assessments for the monozygotic twins, even when reared apart: it is almost as high as the correlation between two successive testings for the same individuals. On the other hand, with school attainments the correlations are much lower for twins reared separately than for twins reared together in the same home.

Several of our critics—Heim and Maddox, for example—have cited the account of analogous cases (described by Newman and others) as proving that intelligence is dependent on environment. Thus, to take an oft-quoted pair, "Helen," who had been trained as a teacher, scored with the Stanford-Binet tests an I.Q. of 116; whereas her twin sister, "Gladys," brought up for much of her childhood in an isolated district of the Canadian Rockies, scored only 92. But, says Newman, her score

> . . . was higher than we might expect considering her scant education; and . . . it seems certain that the great deficiency in education had inhibited the development of the rather high grade of mental ability with which she was *endowed by heredity* (*25*, pp. 136–144).

Thus Newman's interpretation in no way conflicts with ours.

It is sometimes alleged (24) that, since twins are born at the same time, the intrauterine environment must have been the same for both before birth, even if later on their environments differ widely, and that quite conceivably it is the former that is crucial. As it happens, however, this rather gratuitous assumption reverses the actual facts. Embryological and obstetric records

TABLE 1

Correlations for Mental, Educational, and Physical Characteristics

	A. Burt *et al.*						B. Newman *et al.*		
	Mono-zygotic twins reared together	Mono-zygotic twins reared apart	Dizygotic twins reared together	Siblings reared together	Siblings reared apart	Unrelated children reared together	Mono-zygotic twins reared together	Mono-zygotic twins reared apart	Dizygotic twins reared together
Number of pairs*	95	53	127	264	151	136	50	19	51
Intelligence									
Group test	0.944	0.771	0.552	0.545	0.412	0.281	0.922	0.727	0.621
Individual test	0.918	0.863	0.527	0.498	0.423	0.252	0.881	0.767	0.631
Final assessment	0.925	0.874	0.534	0.531	0.438	0.267	—	—	—
Educational									
Reading and spelling	0.951	0.597	0.919	0.842	0.490	0.545	—	—	—
Arithmetic	0.862	0.705	0.748	0.754	0.563	0.478	—	—	—
General attainments	0.983	0.623	0.831	0.803	0.526	0.537	0.892	0.583	0.696
Physical									
Height	0.962	0.943	0.472	0.501	0.536	−0.069	0.932	0.969	0.645
Weight	0.929	0.884	0.586	0.568	0.427	0.243	0.917	0.886	0.631
Head length	0.961	0.958	0.495	0.481	0.506	0.110	0.910	0.917	0.691
Head breath	0.977	0.960	0.541	0.510	0.492	0.082	0.908	0.880	0.654
Eye color	1.000	1.000	0.516	0.554	0.524	0.104	—	—	—

* Figures for boys and girls have been calculated separately and then averaged. In columns 3, 4, 5, and 6 the correlations for head-length, head-breadth, and eye color were based on samples of 100 only.

show that, particularly with twins developed from split ova, the position of each in the uterus, and the subsequent development, is liable to differ widely (*12*, pp. 123ff. and refs.).

I think, therefore, that it may be safely said that, apart from the influence of some preconceived theory, few psychologists nowadays would be inclined to contest the mere fact of mental inheritance: the most that can be plausibly alleged is that its influence is comparatively slight and distinctly elusive.

The Mode of Inheritance. The majority of those who still question the importance of mental inheritance, and many of those who support it, seem by preference to adopt entirely antiquated notions of the way in which inheritable characteristics are transmitted. If, as I have maintained, mental capacities are dependent on the physical characteristics of the brain (or, to speak a little more precisely, on the structural and biochemical qualities of the nervous system), then we should expect those capacities to be inherited in accordance with the same principles that govern the inheritance of physical characteristics; and these principles (except for obscure and apparently exceptional instances of extranuclear heredity) are essentially those commonly associated with the name of Mendel. Many British psychologists, however, feel a strong and not unreasonable prejudice against applying "atomistic theories like Mendel's" to explain the facts of mental life, and consequently, so far as they admit the possibilities of mental inheritance at all, still cling to the old Darwinian principle of blended inheritance. On this view heredity means "the tendency of like to beget like" (the definition quoted by one of them from the *Oxford English Dictionary*). As a result, they commonly assume that the arguments for inheritance must consist in demonstrating resemblances between the parent and his children by means of correlations. When the two parents differ, then the child is still expected to consist in an intermediate blend of both.

The approach of the modern geneticist is the reverse of all this. As he views it, the real problem is rather to explain why in so many instances "like begets unlike." Both for the environmentalists and for the believer in blended inheritance, one of the most puzzling phenomena is the appearance, not merely of extremely dull children, in the families of the well-to-do professional classes, but also of extremely bright children in families where both the cutural and the economic conditions of the parents would, one might imagine, condemn every child to hopeless failure. With the Mendelian hypothesis these anomalies are just what we should anticipate. However, the few critics who are familiar with the Mendelian explanation appear, as a rule, to suppose that it can apply only to discontinuous variations; and point out that intelligence, like stature, exhibits not discontinuous but continuous or graded variation. Hence, so they contend (sometimes citing the experiments of De Vries on "pure lines"), the apparent differences in intelligence between one individual and another must be due almost entirely to differences in environmental conditions.

Mendel himself was the first to indicate how his theory could be extended to account for this particular difficulty. When supplementing his experiments on the hybridization of peas by hybridizing beans, and (as before) crossing white flowered plants with purple, he found, that, whereas with peas the two types always sorted out with no hint of any intermediate color, with beans the offspring displayed "a whole range of hues from white to deep purple." This, he suggested, might be explained by postulating that with beans the color was determined, not by a single pair of alternative factors, but by a *number* of such pairs, each positive factor, when present, contributing a small additional amount of color. And if, as before, the recombinations are the effects of chance unions, then the resulting frequencies would obviously approximate to those of the normal curve.

However, in our early surveys of London children (*5*, *6*), we found that, when complete age groups were tested, the distribution of intelligence departed significantly from that of a perfect normal curve: there was a swollen tail at the lower end, due to an excess of mental defectives, and a smaller enlargement at the upper end. This and other considerations led me to put forward the tentative hypothesis that innate variations in intelligence are due partly to unifactor and partly to multifactor inheritance: i.e., they result from Mendelian factors of two main kinds (no doubt overlapping), viz., (a) major genes responsible for comparatively *large* deviations, usually of an abnormal type, and (b) multiple genes whose effects are *small*, *similar*, and *cumulative*.

Karl Pearson (*26*) endeavoured to test the Mendelian theory in its multifactorial form by comparing its implications with actual figures obtained for height, arm length, and similar physical measurements, collected by himself and Alice Lee, from over 2,000 students and their relatives. The expected correlations which he deduced for various degrees of kinship were in every case far smaller than the observed coefficients. He therefore emphatically rejected the hypothesis of Mendelian inheritance, and fell back on the older theory of blending. However, in deriving his formulae and his expected values, Pearson relied on an over-simplified model. Contrary to what we now know to be the case, he assumed that the effect of assortative mating—the tendency of like to marry like—could be ignored as negligible, and that dominance would in every case be perfect. Ronald Fisher has since undertaken the rather formidable task of deducing more appropriate formulae, which allow for these and other complicating factors (*17*). And with these refinements the calculated correlations fit Pearson's own figures as well as could be wished.

My colleagues and I have applied Fisher's methods (suitably modified) to assessments for intelligence (*12*). The data were secured in the course of surveys of the entire school population in a representative London borough, and covered nearly 1,000 pairs of siblings, together with ratings for parents, and (so far as they were accessible) grandparents, uncles and aunts, and first cousins. The final assessments for the children were obtained by submitting the marks from the group tests to the judgment of the teachers who knew the children best: where the teacher disagreed with the verdict of the marks, the child was interviewed personally, and subjected to further tests, often on several successive occasions. The assessments for the adult members of the family were naturally far less accurate. Nevertheless, in almost every case the correlations computed from the actual data agreed with the theoretical values deduced from the multifactorial hypothesis far better than with the values deduced from any other hypothesis hitherto put forward. The only appreciable discrepancy occurred in the case of first cousins. Here, as for stature, the observed correlation for intelligence is larger than the theoretical; but the difference is not statistically significant, and could readily be explained if (as has been suggested above) variations in intelligence are affected by a few major genes as well as by numerous minor genes. I may add that on sorting out figures for cousins of maternal, paternal, and mixed kinship there is also some slight evidence suggestive of sex linkage.

The Relative Influence of Heredity and Environment. In practical work, however, the question most frequently raised is, not whether differences in intelligence are inherited, nor even how they are inherited, but rather what is the relative influence of heredity as compared with environment. To an omnibus inquiry like this there can be no single answer. We can only try to determine, for this or that type of environment, for this or that population, and for this or that type of assessment, how far the observable results appear to be influenced by each of the two main groups of factors.

As Fisher's analysis has shown, formulae analogous to those used to deduce the expected correlations from the theoretical variances can also be devised for deducing the amount of the constituent variances from the observed correlations. I have ventured to amplify Fisher's methods (mainly on the lines of later work by Mather and Sewall Wright) so as to allow for unreliability and for the systematic effects of environment, i.e., of environmental influences which are correlated with those of heredity, as well as for random effects. The genetic contribution may be regarded as comprising two distinguishable portions: that due to the "fixable" component (or, as Fisher expresses it, to the "essential genotypes") and that due to the "nonfixable" (i.e., deviations due to dominance and similar influences). The data analysed consist of (a) the marks obtained from intelligence tests of the ordinary type taken just as they stand and (b) adjusted assessments obtained by the supplementary methods already described (*12*).

From Table 2 it will be seen that, with the crude test results, taken just as they stand, nearly 23% of the total variance appears due to nongenetic influences, i.e., to environment or to unreliability, and about 77% to genetic factors; with the adjusted assessments only about 12% (or slightly more) is apparently due to nongenetic influences and 88% to genetic factors. This of course means that the common practice of relying on tests alone—usually a group test applied once only—is by no means a satisfactory method of assessing a child's innate ability. Better assessments are obtained by submitting the test scores to the teachers for criticism or correction, and where necessary adjusting them by the methods described above. But such intensive inquiries would be too costly for routine use except in borderline cases.

TABLE 2

Analysis of Variance for Assessments of Intelligence

Source	Unadjusted test scores	Adjusted assessments
Genetic component:		
fixable	40.51	47.92
nonfixable	16.65	21.73
Assortative mating	19.90	17.91
Environment:		
systematic	10.60	1.43
random	5.91	5.77
Unreliability	6.43	5.24
TOTAL	100.00	100.00

Environment appears to influence the test results chiefly in three ways: (a) the cultural amenities of the home and the educational opportunities provided by the school can undoubtedly affect a child's performance in intelligence tests of the ordinary type, since so often they demand an acquired facility with abstract and verbal modes of expression; (b) quite apart from what the child may learn, the constant presence of an intellectual background may stimulate (or seem to stimulate) his latent powers by inculcating a keener motivation, a stronger interest in intellectual things, and a habit of accurate, speedy, and diligent work; (c) in a few rare cases illness or malnutrition during the prenatal or early postnatal stages may, almost from the very start, permanently impair the development of the child's central nervous system. The adjusted assessments may do much towards eliminating the irrelevant effects of the first two conditions; but it is doubtful whether they can adequately allow for the last.

LIMITATIONS INVOLVED IN THESE CONCLUSIONS

As in almost all scientific investigations, the hypothetical model which has formed the basis of our inquiry involves of necessity certain minor simplifications. In particular we have assumed, for purposes of calculation, a sharp distinction between the "major genes" of unifactor inheritance and the "polygenes" of multifactor inheritance, and have treated the latter as contributing equal and additive doses to the sum total of each child's innate intelligence. We have then supposed that the effects of the former would on the whole be excluded if the few obviously pathological cases (mostly found in special schools or institutions) were omitted from our final calculations. However, these assumptions have led several critics to accuse us of

> . . . disrupting the individual personality into atomic bits and discrete pieces which have subsequently to be joined together like a mosaic. . . . Personality [it is argued] is not a mosaic but a seamless whole, and hence the entire Mendelian hypothesis with its particulate genes, each producing a unit-character or adding another unit to the same character, is quite inapplicable to the facts of conscious behavior, and therefore to the study of mental capacity.

Objections of this kind could of course be used just as well to prove that a neuronic theory of the central nervous system is incompatible with the facts of conscious behavior or of individual variations in ability, since nerve cells or "neural bonds" are equally "particulate." Nevertheless, even those neurologists who prefer to start from a "field theory" do not wholly reject the neuronic hypothesis (cf. *30*). In both cases the difficulties raised owe their force chiefly to the fact that there is a vast series of elusive processes intervening between theory and observable results which the critic is exceedingly apt to forget. Moreover, criticisms like those just cited plainly rest on an obsolete version of the Mendelian doctrine. No geneticist today, I imagine, accepts the hypothesis of the autonomous corpuscular gene; and the genotypic endowment of the individual can only affect the phenotypic resultant through the mediation of innumerable obscure biochemical steps.

In our original papers Howard and I tried carefully to guard against recurrent objections of this type. As we pointed out, the phrase "multiple factors" may be used to cover either (a) relatively numerous loci each with only two allelomorphs or (b) a single locus (or relatively few loci) each with numerous allelomorphs, or possibly (c) some combination of the two. Hypothesis b by itself would hardly seem to fit the facts. We are inclined to think that factors of all the various types may be operative in varying degrees, and that the attempt to classify factors or genes should not be too severely pressed.

We further assumed that in all probability the influence of such factors on the individual's observable intelligence was mainly the indirect result of their influence on the development of his central nervous system, and was presumably effected by modifying growth rates. And we expressly stated that the ultimate influence of any one "gene" upon intelligence might be but one of its multifarious consequences, and possibly a comparatively remote consequence at that. Some genes may have a larger share in the final result; others a smaller; and the rough classification of genes into "major" genes and "minor" was adopted primarily with a view to simplifying our general discussion. We ourselves find the "theory of chromosomal hierarchy," advanced by Goldschmidt (*22*), especially attractive as a basis for the ultimate hierarchical differentiation of mental ability.

However, this is not the place to enlarge on these speculative interpretations. We fully admit that the simplifications involved in our hypothetical model mean that the figures finally deduced can be no more than approximations. But we maintain that the error of approximation, however large, will nevertheless be smaller than the amount of "unreliability" inevitably involved in all such measurements.

In any case we must repeat that the conclusions reached are at best only valid in reference to the particular conditions under which they were obtained. They would not necessarily hold good (a) of other mental traits, (b) of different methods of assessment, (c) of a population at a different cultural level: much less would they hold good if there were any subsequent change (d) in the present distribution of environmental and genetic characteristics, or (e) in the influences affecting their mutual interaction.

THE INHERITANCE OF OTHER MENTAL QUALITIES

The mental trait in which we have been chiefly interested is the general factor of "intelligence." This, we readily concede, is but one psychological ingredient in ordinary everyday behavior. Our critics are accustomed to reproach us for apparently ignoring the rest. There are several obvious reasons why we have dealt less fully with other determinants. The first and the most conclusive is that comparatively little evidence is as yet available in regard to the genetic composition of other mental factors. We have, however, in the course of our surveys, met with an appreciable number of cases which strongly suggest the inheritability of certain specific abilities and disabilities, or rather of certain ill-defined tendencies that presumably underlie such disabilities—particularly in our studies of memory, of visual, auditory, and kinaesthetic imagery, and of verbal, numerical, artistic, and musical aptitudes (5). We have encountered other cases which suggest the inheritability of general emotional instability, of temperamental qualities like introversion and extraversion, and of certain quasi-instinctive tendencies like sex and bad temper (or "pugnacity," as McDougall would have termed it). But emotional tendencies are always much more liable to be influenced and altered by postnatal experiences, and in any case undergo considerable changes during the developmental stages, so that all attempts at assessment are apt to be precarious. The most convincing evidence is afforded by the family histories of children brought up in orphanages or residential institutions (7, especially p. 14 and refs.). In all these instances the evidence, such as it is, appears to indicate that each of the characteristics mentioned, cognitive and temperamental alike, are influenced both by unifactor and by multifactor modes of inheritance—the latter tending on the whole to predominate, but in varying degrees. This after all is what is commonly found in investigations on physical characteristics. Characteristics depending mainly (though not entirely) on the development of the long bones—such as arm length, leg length, stature, and perhaps the longer dimensions of the face and head—yield correlations of much the same magnitude as those obtained for general intelligence. Others, like cephalic index, breadth at hips, and particularly facial characteristics, seem to be rather more subject to the influence of single (or relatively few) genes, which often appear to exhibit almost perfect dominance. Variations in the cartilaginous features, and particularly those depending on flesh and fat, are far more readily affected by environmental influences (for data, see *34* and refs.).

But secondly there is a practical reason for concentrating attention first of all, and chiefly, on the inheritability of general intelligence. General intelligence, as we have seen, accounts for four times as much of the variance as any other identifiable factor. Galton's pronouncement is thus fully borne out by our own results:

> . . . general ability appears far more important than special gifts; and, where the allowance granted by nature is inadequate, the keenest will and the stoutest industry must strive in vain.

Moreover, the child's innate endowment of intelligence sets an upper limit to the best he can possibly attain. No one would expect a mongolian imbecile, with the most skilful coaching in the world, to achieve the scholastic knowledge of an average child. In the same way, no one should expect a child who is innately dull to gain a scholarship to a grammar school, or one whose inborn ability is merely average to win first class honors at Oxford or Cambridge. No doubt, in any individual case the ascertainment of this upper limit can never be a matter of absolute certainty: an I.Q. derived from tests alone falls far short of a trustworthy indication. Hence education authorities, like life insurance companies, have to follow Butler's maxim, and take probability for their guide. They cannot, however, afford to risk a lavish expenditure on cases where there are fairly heavy odds against success.

Nevertheless, it must be owned that, in Britain at any rate, the existing administrative machinery is far less efficient than it might be. In an inquiry made just before the introduction of the Education Act of 1944 we found that "approximately 40 per cent of those whose innate abilities are of university standard are failing to reach the university" (*9*). The new arrangements proposed by the act, and the postwar changes in the economic circumstances of the various social classes, have appreciably altered the entries to "grammar schools"; but it is too early to say how far they have affected the composition of the universities. For grammar schools the entries vary considerably from one area to another and even from year to year. The general aim is apparently to provide a grammar school education for the brightest 20% of each age group; this corresponds to a borderline of about 113 IQ. On this basis the round figures in Table 3 represent estimates deduced from the information so far available.

TABLE 3

Estimates for Grammar School Entries

	General population*	Proportions over +0.85 SD	Expected entries	Actual entries
Middle Class	20%	.40	.08 (40%)	55%
Manual Class	80%	.15	.12 (60%)	45%
TOTAL	100%		.20 (100%)	100%

* The *Census of Great Britain* (1951) gives an estimate of 81.7 for males over 15, occupied or retired, falling into the classes of "Skilled," "Mixed," and "Unskilled" workers; these I have pooled in a single broad group.

A more intensive survey of two contrasted areas ("industrial" and "prosperous") has just been published by J. E. Floud and her colleagues. Briefly their conclusion is that, in both areas, gross material handicaps have been greatly lessened, though not entirely eliminated. Nevertheless, in their view, the "problem of social wastage" is by no means overcome. The fact that, even today, a large proportion of the children from the working classes fail to reach the higher levels of scholastic instruction

Floud herself attributes to the way in which the "educational ladder" is still widely regarded as a "middle class prerogative" to be anxiously watched and jealously preserved (*20*).

I myself should be inclined to look rather to a difference in the aims and aspirations which are traditional in the different classes. The consequences seemed clearly revealed by the differences in the after-careers of LCC scholarship winners from the lower and the middle classes respectively whom my colleagues and I have been able to follow up (cf. *5*, *9*, and refs.). The figures show that the abler children from the working classes, even when they have obtained free places or scholarships at secondary schools of the "grammar" type, frequently fail to stay the course; by the time they are sixteen the attractions of high wages and of cheap entertainment during leisure hours prove stronger than their desire for further knowledge and skill, and easily overcome their original resolve to face a long prospect of hard sedentary work *in statu pupillari*. As a headmaster of a secondary school which receives both types of boy has put it:

> The working class parent wants his boy to be "selected" chiefly to prove that "his child is as good as anybody else's"; having done so, he will withdraw him at the earliest possible opportunity: I do not press them, for the presence of those whose chief interest in life is television, Hollywood films, football pools, "the dogs," and "the girls" has not improved the tone of my school. And too often these "ignobler attitudes and crude ambitions" are shared by, and encouraged by, both parents and friends.

On the other hand, if I may quote the report of an experienced university teacher:

> That section of the middle class which seeks, by paying fees that it can ill afford, to assist its children to climb, *via* a "public school" or an independent "grammar school," to a University education and a good honours degree in the humaner subjects, is animated by a traditional morale which is comparatively rare in children and parents from the other classes: those who nowadays come here on grants, at no cost to themselves or their parents, are, on the whole, most irregular attendants, and the least satisfactory students.

Or, to quote a still more recent pronouncement of the High Master of St. Paul's:[1]

> The parents are themselves imbued with four traditional ideals which they hand on to their posterity: self-discipline, a community spirit, the Christian religion, and a readiness to accept social responsibility even at the sacrifice of material enjoyments—a genuine *noblesse oblige.*

Underlying all these differences in outlook I myself am tempted to suspect an innate and transmissible difference in temperamental stability and in character, or in the neurophysiological basis on which such temperamental and moral differences tend to be built up. Tradition may explain much: it can hardly account for all. However, it would be idle to pursue such speculations here in the absence of more adequate data.

NEED FOR FURTHER RESEARCH

To my mind the most pressing need at the moment is for more extensive research. Hitherto the most active investigators have been research-students, with little or no experience of the ways of children or the conditions of the classroom. For the requisite facilities, and for any supplementary information they may want, they have to rely on the goodwill of the busy teacher. And one important factor is still more often overlooked: the motivation of the children themselves. This takes two obvious forms. There is first what may be called direct or short distance motivation. As actual trial has repeatedly shown, neither pupils nor students are likely to exert their full powers in a test conducted merely in the interests of someone else's research; but when the examinees' entrance to a grammar school or university depends on their performances, the average score may rise 5 to 10 IQ points higher. Secondly, there are the effects of indirect or long distance motivation: the influence of parental attitudes, of teachers' exhortations, and (most of all perhaps with British lads) of the social pressure that arises from the opinions and the comments of their school fellows, and from the tone and atmosphere of the school or class to which the pupil belongs. It is therefore hardly surprising if the results obtained in mere academic researches are at times disappointing, and evince a comparatively low reliability.

---

[1] A. N. Gilkes, *Independence in Education,* 1957.

I myself had the good fortune to be appointed a member of the school inspectorate, and so not only acquired a firsthand knowledge of the schools, teachers, and pupils, but enjoyed full authority to interrupt the timetable, examine private records, and requisition whatever information might be wanted. Here therefore I must again record my indebtedness both to the London County Council for facilitating these inquiries and financing their publication, and to the teachers, social workers, and school medical officers, who rendered such wholehearted cooperation in all our investigations. It is to be hoped that other education authorities will in the near future perceive the practical value of systematically organized inquiries of this kind, conducted on an official basis by those who know the schools from inside.

The basic researches must be carried out on children rather than on students or adults. With adults, it is much harder to achieve accurate assessments; innate tendencies are already obscured and overlaid; and detailed family histories difficult to obtain. In allocating adults to appropriate occupations, whether in civil life or in the fighting services, innate capacities are of little consequence, and accordingly tend to be ignored. As a result the severest critics of researches on heredity are generally those whose experience of mental testing has been gained in the adult field. This was manifest during the symposium recently arranged by the Psychological Section of the British Association. Several contributors, for instance, echoed remarks like those of H. E. Jones that "potential investigators should now be advised that the nature-nurture controversy has been shown to be an unproductive field of research" (*14*); and similar conclusions have more recently been voiced by writers like Maddox, Renshaw, and others, who deplore what they call "the resurrection of the nature-nurture controversy" as "wholly anachronistic" (*24, 28*).

But whether or not this is now the prevailing view among psychologists, it is not the opinion of leading geneticists. Goldschmidt, Snyder, Calvin Hall, and Calvin Stone all deplore the neglect of genetic inquiries by psychologists of the present day.

Goldschmidt in particular has criticized "the extreme belief of contemporary psychologists in the power of environment"—a belief which he ascribes to the "doctrine of universal conditioning, which has, until lately, dominated the behaviourist school" (*16*). In this country eminent authorities like R. A. Fisher (*18*, *19*) and Fraser Roberts (*29*) have not only urged the importance of genetic studies for psychology, but have made noteworthy contributions of their own.

Mere statistical inquiries, however, can never suffice. In this country one of the most obvious needs is a series of intensive studies of able boys and girls similar to those already carried out on the backward, the defective, the neurotic, and the delinquent. The brilliant investigation of 1,000 gifted children, started by Terman and his colleagues over 30 years ago, and so assiduously followed up, furnishes an admirable model (*35*, *36*). But nothing like conclusive answers can be given to the questions here raised until far more extensive research has been undertaken on the fundamental problems of psychogenetics. An adequate understanding of the basic processes can be secured only if we start with carefully planned experiments on lowlier creatures, where pure strains can be obtained, breeding controlled, and successive generations more speedily raised. When we know more about the genetics of intelligence in animals, we may be able to construct with greater confidence a more exact hypothesis regarding the transmission of intelligence in man.

SOCIAL IMPLICATIONS

After all, the practical importance of the issues involved is so profound and so far-reaching that it would be fatal to dismiss the problem as "unproductive" or "anachronistic" without attempting to settle it one way or the other. No democratic state can afford to pass it by. If in the end the views of the early pioneers should turn out to have been approximately correct—if "innate intelligence varies between limits at least as wide as IQ's of 50 and 150," and if too "the average intelligence of the general population at maturity is little if at all above that of an average child of thirteen or thirteen-and-a-half"—then the bearing of genetic variation on the national and social questions of the present day would be all too obvious.

In our laudable eagerness to improve the lot of our own generation, we have been tempted to close our eyes to the effects of present policy on the generations to come. The over-all efficiency of the citizens who make up a nation or a state must in the last resort depend on what has been called its "chromosomal pool." Improved environmental amenities can of themselves ensure no lasting results; but the changes in a nation's genetic constitution are likely to prove irreversible. Throughout almost the entire Western world, and to a less extent in other areas as well, the last half-century or so has witnessed radical modifications in the conditions that previously governed marriage and fertility. Increased mobility, enhanced freedom, new means of production, new methods of government, the progressive reduction of the traditional barriers separating different peoples and different social groups, and, in our own country, the recent extensions of the educational ladder, all these transformations are visibly disturbing the stability alike of classes and of races. Almost inevitably they must alter the genetic constitution of what have hitherto been the dominant nations and the dominant stocks within those nations.

From history and from past experience we know full well how changes in the balance between inbreeding and outbreeding can affect the later destinies of populations, remoulding them sometimes for good, sometimes for ill. Geneticists have repeatedly drawn attention to the processes at work, and indicated the need for studying the possible effects of contemporary trends (cf., e.g., *15*, pp. 49ff. and *19*, pp. 170ff.). Yet current sociological textbooks seem almost wholly to ignore them. Surely it is high time that the social psychologist should now take up the question and plan a systematic series of investigations on the intricate but highly important issues involved. May I therefore conclude by endorsing, in the strongest possible terms, the verdict of Stone (*16*), quoted by Bingham in the last letter he wrote to me:

> It is a matter of shame and regret that only an amateurish beginning has been made by psychologists in applying modern genetic methods to fundamental study in the nature-nurture area.

And may we hope that in the near future Bingham's "deepest wish" may be fulfilled, and that "a small band of enthusiasts will come forward to explore afresh this urgent and many-sided field of research."

REFERENCES

1. BAGLEY, W. C. *The educational process.* New York: Macmillan, 1912.
2. BURKS, BARBARA. *27th Yearbook National Society for the Study of Education*, 1928, 1, 219.
3. BURT, C. Experimental tests of general intelligence. *Brit. J. Psychol.*, 1909, 3, 94–177.
4. BURT, C. The inheritance of mental characteristics. *Eugen. Rev.*, 1912, 4, 168–200.
5. BURT, C. *Annual reports of the psychologist to the London County Council.* London: London County Council, 1914–1931.
6. BURT, C. *The distribution and relations of educational abilities.* London: P. S. King & Sons, 1917.
7. BURT, C. Is the doctrine of instincts dead? *Brit. J. educ. Psychol.*, 1941, 11, 155–172; 1943, 13, 1–16.
8. BURT, C. The education of the adolescent: The psychological implications of the Norwood Report. *Brit. J. educ. Psychol.*, 1943, 13, 126–140.
9. BURT, C. Ability and income. *Brit. J. educ. Psychol.*, 1943, 13, 83–98.
10. BURT, C. The differentiation of intellectual ability. *Brit. J. educ. Psychol.*, 1954, 24, 76–90.
11. BURT, C. The evidence for the concept of intelligence. *Brit. J. educ. Psychol.*, 1955, 25, 158–177.
12. BURT, C., & HOWARD, M. The multifactorial theory of inheritance and its application to intelligence. *Brit. J. statist. Psychol.*, 1956, 9, 95–131.
13. CAMPBELL, F. *Eleven plus and all that.* London: Watts, 1956.
14. CARMICHAEL, L. (Ed.) *Manual of child psychology.* New York: Wiley, 1946.
15. DARLINGTON, C. D., & MATHER, K. *Genes, plants, and people.* London: Allen & Unwin, 1950.
16. DUNN, L. C. (Ed.) *Eugenics in the twentieth century.* New York: Macmillan, 1921.

17. Fisher, R. A. Correlation between relatives on the supposition of Mendelian inheritance. *Trans. roy. Soc., Edin.*, 1918, 52, 399–433.
18. Fisher, R. A. The causes of human variability. *Eugen. Rev.*, 1919, 11, 213–220.
19. Fisher, R. A. *The genetical theory of natural selection*. Oxford: Clarendon Press, 1950.
20. Floud, J. E., Halsey, A. H., & Martin, F. M. *Social class and educational opportunity*. London: Heinemann, 1956.
21. Galton, F. *Hereditary genius*. London: Macmillan, 1869.
22. Goldschmidt, R. B. *Theoretical genetics*. Univ. California Press, 1955.
23. Lashley, K. S., & Clark, G. The cytoarchitecture of the cerebral cortex of *Ateles*: A critical examination of architectonic studies. *J. comp. Neurol.*, 1946, 85, 223–305.
24. Maddox, H. Nature-nurture balance sheets. *Brit. J. educ. Psychol.*, 1957, 27, 166–175.
25. Newman, H. H. Twins and super-twins. London: Hutchinson, 1942.
26. Pearson, K. On a generalized theory of alternative inheritance with special reference to Mendel's laws. *Phil. Trans.*, 1904, 203, 53–87.
27. Piaget, J. *The psychology of intelligence*. London: Kegan Paul, 1950.
28. Renshaw, T. Burt's concept of intelligence. *Brit. Psychol. Soc. Bull.*, 1957.
29. Roberts, J. A. F. *An introduction to medical genetics*. Oxford: Oxford Univer. Press, 1940.
30. Sholl, D. A. *The organization of the cerebral cortex*. London: Methuen, 1956.
31. Spearman, C. *Abilities of man*. London: Macmillan, 1927.
32. Spencer, H. *Principles of psychology*. London: William & Norgate, 1870.
33. Stone, C. P. Methodological resources for the experimental study of innate behavior as related to environmental factors. *Psychol. Rev.*, 1947, 54, 342–347.
34. Tanner, J. M. The inheritance of morphological and physiological traits. *Ap.* Sorsby, A. *Clinical genetics*. London: Butterworth, 1953.
35. Terman, L. M. (Ed.) *Genetic studies of genius*. Stanford: Stanford Univer. Press, 1925, 1947.
36. Terman, L. M. The discovery and encouragement of exceptional talent. *Amer. Psychologist*, 1954, 9, 221–230.

37. Thurstone, L. L. *The vectors of mind*. Chicago: Univer. Chicago Press, 1935.
38. Watson, J. B. *Behaviourism*. London: Paul, Trench Trubner, 1931.

15

Heredity, Environment, and the Question "How"

ANNE ANASTASI

It was noted in the introduction to this group of papers on the nature-nurture problem that dogmatic "either-or" formulations have been replaced by a study of the ways in which hereditary and environmental factors interact in the process of development. The final paper in this group shows by explanation and examples what this way of thinking involves. By Anne Anastasi, it is entitled "Heredity, Environment, and the Question 'How'," Psychological Review, *1958, 65, 197–208. It is reprinted with the permission of the author and the American Psychological Association.*

Two or three decades ago, the so-called heredity-environment question was the center of lively controversy. Today, on the other hand, many psychologists look upon it as a dead issue. It is now generally conceded that both hereditary and environmental factors enter into all behavior. The reacting organism is a product of its genes and its past environment, while present environment provides the immediate stimulus for current behavior. To be sure, it can be argued that, although a given trait may result from the combined influence of hereditary and environmental factors, a specific difference in this trait between individuals or between groups may be traceable to either hereditary or environmental factors alone. The design of most traditional investigations undertaken to identify such factors, however, has been such as to yield inconclusive answers. The same set of data has frequently led to opposite conclusions in the hands of psychologists with different orientations.

Nor have efforts to determine the proportional contribution of hereditary and environmental factors to observed individual differences in given traits met with any greater success. Apart from difficulties in controlling conditions, such investigations have usually been based upon the implicit assumption that hereditary and environmental factors combine in an additive fashion. Both geneticists and psychologists have repeatedly demonstrated, however, that a more tenable hypothesis is that of interaction (*15*, *22*, *28*, *40*). In other words, the nature and extent of the influence of each type of factor depend upon the contribution of the other. Thus the proportional contribution of heredity to the variance of a given trait, rather than being a constant, will vary under different environmental conditions. Similarly, under different hereditary conditions, the relative contribution of environment will differ. Studies designed to estimate the proportional contribution of heredity and environment, however, have rarely included measures of such interaction. The only possible conclusion from such research would thus seem to be that both heredity and environment contribute to all behavior traits and that the extent of their respective contributions cannot be specified for any trait. Small wonder that some psychologists regard the heredity-environment question as unworthy of further consideration!

But is this really all we can find out about the operation of heredity and environment in the etiology of behavior? Perhaps we have simply been asking the wrong questions. The traditional questions about heredity and environment may be intrinsically unanswerable. Psychologists began by asking *which* type of factor, hereditary or environmental, is responsible for individual differences in a given trait. Later, they tried to discover *how much* of the variance was attributable to heredity and how much to environment. It is the primary contention of this paper that a more fruitful approach is to be found in the question "*How?*" There is still much to be learned about the specific *modus operandi* of hereditary and environmental factors in the development of behavioral differences. And there are several current lines of research which offer promising techniques for answering the question "How?"

VARIETY OF INTERACTION MECHANISMS

Hereditary Factors

If we examine some of the specific ways in which hereditary factors may influence behavior, we cannot fail but be impressed by their wide diversity. At one extreme, we find such conditions as phenylpyruvic amentia and amaurotic idiocy. In these cases, certain essential physical prerequisites for normal intellectual development are lacking as a result of hereditary metabolic disorders. . . .

A somewhat different situation is illustrated by hereditary deafness, which may lead to intellectual retardation through interference with normal social interaction, language development, and schooling. In such a case, however, the hereditary handicap can be offset by appropriate adaptations of training procedures. It has been said, in fact, that the degree of intellectual backwardness of the deaf is an index of the state of development of special instructional facilities. As the latter improve, the intellectual retardation associated with deafness is correspondingly reduced.

A third example is provided by inherited susceptibilty to certain physical diseases, with consequent protracted ill health. If environmental conditions are such that illness does in fact develop, a number of different behavioral effects may follow. Intellectually, the individual may be handicapped by his inability to attend school regularly. On the other hand, depending upon age of onset, home conditions, parental status, and similar factors, poor health may have the effect of concentrating the individual's energies upon intellectual pursuits. The curtailment of participation in athletics and social functions may serve to strengthen interest in reading and other sedentary activities. Concomitant circumstances would also determine the influence of such illness upon personality development. And it is well known that the latter effects could run the gamut from a deepening of human sympathy to psychiatric breakdown.

Finally, heredity may influence behavior through the mechanism of social stereotypes. A wide variety of inherited physical characteristics have served as the visible cues for identifying such stereotypes. These cues thus lead to behavioral restrictions or opportunities and—at a more subtle level—to social attitudes and expectancies. The individual's own self concept tends gradually to reflect such expectancies. All of these influences eventually leave their mark upon his abilities and inabilities, his emotional reactions, goals, ambitions, and outlook on life.

The geneticist Dobzhansky illustrates this type of mechanism by means of a dramatic hypothetical situation. He points out that, if there were a culture in which the carriers of blood group AB were considered aristocrats and those of blood group O laborers, then the blood-group genes would become important hereditary determiners of behavior (*12*, p. 147). Obviously the association between blood group and behavior would be specific to that culture. But such specificity is an essential property of the causal mechanism under consideration.

More realistic examples are not hard to find. The most familiar instances occur in connection with constitutional types, sex, and race. Sex and skin pigmentation obviously depend upon heredity. General body build is strongly influenced by hereditary components, although also susceptible to environmental modification. That all these physical characteristics may exert a pronounced effect upon behavior within a given culture is well known. It is equally apparent, of course, that in different cultures the behavioral correlates of such hereditary physical traits may be quite unlike. A specific physical cue may be completely unrelated to individual differences in psychological traits in one culture, while closely correlated with them in another. Or it may be associated with totally dissimilar behavior characteristics in two different cultures.

It might be objected that some of the illustrations which have been cited do not properly exemplify the operation of hereditary mechanisms in behavior development, since hereditary factors enter only indirectly into the behavior in question. Closer examination, however, shows this distinction to be untenable. First it may be noted that the influence of heredity upon behavior is always indirect. No psychological trait is ever inherited as such. All we can ever say directly from behavioral observations is that a given trait shows evidence of being influenced by certain "inheritable unknowns." This merely defines a problem for genetic research; it does not provide a causal explanation. Unlike the blood groups, which are close to the level of primary gene products, psychological traits are related to genes by highly indirect and devious routes. Even the mental deficiency associated with phenylketonuria is several steps removed from the chemically defective genes that represent its hereditary basis. Moreover, hereditary influences cannot be dichotomized into the more direct and the less direct. Rather do they represent a whole "continuum of indirectness," along which are found all degrees of remoteness of causal links. The examples already cited illustrate a few of the points on this continuum.

It should be noted that as we proceed along the continuum of indirectness, the range of variation of possible outcomes of hereditary factors expands rapidly. At each step in the causal chain, there is fresh opportunity for interaction with other hereditary factors as well as with environmental factors. And since each interaction in turn determines the direction of subsequent interactions, there is an ever-widening network of possible outcomes. If we visualize a simple sequential grid with only two alternatives at each point, it is obvious that there are two possible outcomes in the one-stage situation, four outcomes at the second stage, eight at the third, and so on in geometric progression. The actual situation is undoubtedly much more complex, since there will usually be more than two alternatives at any one point.

In the case of the blood groups, the relation to specific genes is so close that no other concomitant hereditary or environmental conditions can alter the outcome. If the organism survives at all, it will have the blood group determined by its genes. Among psychological traits, on the other hand, some variation in outcome is always possible as a result of concurrent circumstances. Even in cases of phenylketonuria, intellectual development will exhibit some relationship with the type of care and training available to the individual. That behavioral outcomes show progressive diversification as we proceed along the continuum of indirectness is brought out by the other examples which were cited. Chronic illness *can* lead to scholarly renown or to intellectual immaturity; a mesomorphic physique *can* be a contributing factor in juvenile delinquency or in the attainment of a college presidency! Published data on Sheldon somatotypes provide some support for both of the latter outcomes.

Parenthetically, it may be noted that geneticists have sometimes used the term "norm of reaction" to designate the range of variation of possible outcomes of gene properties (cf. *13*, p. 161). Thus heredity sets the "norm" or limits within which environmental differences determine the eventual outcome. In the case of some traits, such as blood groups or eye color, this norm is much narrower than in the case of other traits. Owing to the rather different psychological connotations of both the words "norm" and "reaction," however, it seems less confusing to speak of the "range of variation" in this context.

A large portion of the continuum of hereditary influences which we have described coincides with the domain of somatopsychological relations, as defined by Barker et al. (*6*). Under this heading, Barker includes "variations in physique that affect the psychological situation of a person by influencing the effectiveness of his body as a tool for actions or by serving as a stimulus to himself or others" (*6*, p. 1). Relatively direct neurological influences on behavior, which have been the traditional concern of physiological psychology, are excluded from this definition, Barker being primarily concerned with what he calls the "social psychology of physique." Of the examples cited in the present paper, deafness, severe illness, and the physical characteristics associated with social stereotypes would meet the specifications of somatopsychological factors.

The somatic factors to which Barker refers, however, are not limited to those of hereditary origin. Bodily conditions attributable to environmental causes operate in the same sorts of somatopsychological relations as those traceable to heredity. In fact, heredity-environment distinctions play a minor part in Barker's approach.

Environmental Factors: Organic

Turning now to an analysis of the role of environmental factors in behavior, we find the same etiological mechanisms which were observed in the case of hereditary factors. First, however, we must differentiate between two classes of environmental influences: (a) those producing organic effects which may in turn influence behavior and (b) those serving as direct stimuli for psychological reactions. The former may be illustrated by food intake or by exposure to bacterial infection; the latter, by tribal initiation ceremonies or by a course in algebra. There are no completely satisfactory names by which to designate these two classes of influences. In an earlier paper by Anastasi and Foley (*4*), the terms "structural" and "functional" were employed. However, "organic" and "behavioral" have the advantage of greater familiarity in this context and may be less open to misinterpretation. Accordingly, these terms will be used in the present paper.

Like hereditary factors, environmental influences of an organic nature can also be ordered along a continuum of indirectness with regard to their relation to behavior. This continuum closely parallels that of hereditary factors. One end is typified by such conditions as mental deficiency resulting from cerebral birth injury or from prenatal nutritional inadequacies. A more indirect etiological mechanism is illustrated by severe motor disorder—as in certain cases of cerebral palsy—*without* accompanying injury to higher neurological centers. In such instances, intellectual retardation may occur as an indirect result of the motor handicap, through the curtailment of educational and social activities. Obviously this causal mechanism corresponds closely to that of hereditary deafness, cited earlier in the paper.

Finally, we may consider an environmental parallel to the previously discussed social stereotypes which were mediated by hereditary physical cues. Let us suppose that a young woman with mousy brown hair becomes transformed into a dazzling golden blonde through environmental techniques currently available in our culture. It is highly probable that this metamorphosis will alter, not only the reactions of her associates toward her, but also her own self concept and subsequent behavior. The effects could range all the way from a rise in social poise to a drop in clerical accuracy!

Among the examples of environmentally determined organic influences which have been described, all but the first two fit Barker's definition of somatopsychological factors. With the exception of birth injuries and nutritional deficiencies, all fall within the social psychology of physique. Nevertheless, the individual factors exhibit wide diversity in their specific *modus operandi*—a diversity which has important practical as well as theoretical implications.

Environmental Factors: Behavioral

The second major class of environmental factors—the behavioral as contrasted to the organic—are by definition direct influences. The immediate effect of such environmental factors is always a behavioral change. To be sure, some of the initial behavioral effects may themselves indirectly affect the individual's later behavior. But this relationship can perhaps be best conceptualized in terms of breadth and permanence of effects. Thus it could be said that we are now dealing, not with a continuum of indirectness, as in the case of hereditary and organic-environmental factors, but rather with a continuum of breadth.

Social class membership may serve as an illustration of a relatively broad, pervasive, and enduring environmental factor. Its influence upon behavior development may operate through many channels. Thus social level may determine the range and nature of intellectual stimulation provided by home and community through books, music, art, play activities, and the like. Even more far-reaching may be the effects upon interests and motivation, as illustrated by the desire to perform abstract intellectual tasks, to surpass others in competitive situations, to succeed in school, or to gain social approval. Emotional and social traits may likewise be influenced by the nature of interpersonal relations characterizing homes at different socio-economic levels. Somewhat more restricted in scope than social class, although still exerting a relatively broad influence, is amount of formal schooling which the individual is able to obtain.

A factor which may be wide or narrow in its effects, depending upon concomitant circumstances, is language handicap. Thus the bilingualism of an adult who moves to a foreign country with inadequate mastery of the new language represents a relatively limited handicap which can be readily overcome in most cases. At most, the difficulty is one of communication. On the other hand, some kinds of bilingualism in childhood may exert a retarding influence upon intellectual development and may under certain conditions affect personality development adversely (*2*, *5*, *10*). A common pattern in the homes of immigrants is that the child speaks one language at home and another in school, so that his knowledge of each language is limited to certain types of situations. Inadequate facility with the language of the school interferes with the acquisition of basic concepts, intellectual skills, and information. The frustration engendered by scholastic difficulties may in turn lead to discouragement and general dislike of school. Such reactions can be found, for example, among a number of Puerto Rican children in New York City schools (*3*). In the case of certain groups, moreover, the child's foreign language background may be perceived by himself and his associates as a symbol of minority group status and may thereby augment any emotional maladjustment arising from such status (*34*).

A highly restricted environmental influence is to be found in the opportunity to acquire specific items of information occurring in a particular intelligence test. The fact that such opportunities may vary with culture, social class, or individual experiential background is at the basis of the test user's concern with the problem of coaching and with "culture-free" or "culture-fair" tests (cf. *1*, *2*). If the advantage or disadvantage which such experiential differences confer upon certain individuals is strictly confined to performance on the given test, it will obviously reduce the validity of the test and should be eliminated.

In this connection, however, it is essential to know the breadth of the environmental influence in question. A fallacy inherent in many attempts to develop culture-fair tests is that the breadth of cultural differentials is not taken into account. Failure to consider breadth of effect likewise characterizes certain discussions of coaching. If, in coaching a student for a college admission test, we can improve his knowledge of verbal concepts and his reading comprehension, he will be better equipped to succeed in college courses. His performance level will thus be raised, not only on the test, but also on the criterion which the test is intended to predict. To try to devise a test which is not susceptible to such coaching would merely reduce the effectiveness of the test. Similarly, efforts to rule out cultural differentials from test items so as to make them equally "fair" to subjects in different social classes or in different cultures may merely limit the usefulness of the test, since the same cultural differentials may operate within the broader area of behavior which the test is designed to sample.

METHODOLOGICAL APPROACHES

The examples considered so far should suffice to highlight the wide variety of ways in which hereditary and environmental factors may interact in the course of behavior development. There is clearly a need for identifying explicitly the etiological mechanism whereby any given hereditary or environmental condition ultimately leads to a behavioral characteristic—in other words, the "how" of heredity and environment. Accordingly, we may now take a quick look at some promising methodological approaches to the question "how."

Within the past decade, an increasing number of studies have been designed to trace the connection between specific factors in the hereditary backgrounds or in the reactional biographies of individuals and their observed behavioral characteristics. There has been a definite shift away from the predominantly descriptive and correlational approach of the earlier decades toward more deliberate attempts to verify explanatory hypotheses. Similarly, the cataloguing of group differences in psychological traits has been giving way gradually to research on *changes* in group characteristics following altered conditions.

Among recent methodological developments, we have chosen seven as being particularly relevant to the analysis of etiological mechanisms. The first represents an extension of selective breeding investigations to permit the identification of specific hereditary conditions underlying the observed behavioral differences. When early selective breeding investigations such as those of Tryon (*36*) on rats indicated that "maze learning ability" was inherited, we were still a long way from knowing what was actually being transmitted by the genes. It was obviously not "maze learning ability" as such. Twenty—or even ten—years ago, some psychologists would have suggested that it was probably general intelligence. And a few might even have drawn a parallel with the inheritance of human intelligence.

But today investigators have been asking: Just what makes one group of rats learn mazes more quickly than the other? Is it differences in motivation, emotionality, speed of running, general activity level? If so, are these behavioral characteristics in turn dependent upon group differences in glandular development, body weight, brain size, biochemical factors, or some other organic conditions? A number of recent and ongoing investigations indicate that attempts are being made to trace, at least part of the way, the steps whereby certain chemical properties of the genes may ultimately lead to specified behavior characteristics.

An example of such a study is provided by Searle's (*31*) follow-up of Tryon's research. Working with the strains of maze-bright and maze-dull rats developed by Tryon, Searle demonstrated that the two strains differed in a number of emotional and motivational factors, rather than in ability. Thus the strain differences were traced one step further, although many links still remain to be found between maze learning and genes. A promising methodological development within the same general area is to be found in the recent research of Hirsch and Tryon (*18*). Utilizing a specially devised technique for measuring individual differences in behavior among lower organisms, these investigators launched a series of studies on selective breeding for behavioral characteristics in the fruit fly, *Drosophila*. Such research can capitalize on the mass of available genetic knowledge regarding the morphology of *Drosophila*, as well as on other advantages of using such an organism in genetic studies.

Further evidence of current interest in the specific hereditary factors which influence behavior is to be found in an extensive research program in progress at the Jackson Memorial Laboratory, under the direction of Scott and Fuller (*30*). In general, the project is concerned with the behavioral characteristics of various breeds and cross-breeds of dogs. Analyses of some of the data gathered to date again suggest that "differences in performance are produced by differences in emotional, motivational, and peripheral processes, and that genetically caused differences in central processes may be either slight or non-existent" (*29*, p. 225). In other parts of the same project, breed differences in physiological characteristics, which may in turn be related to behavioral differences, have been established.

A second line of attack is the exploration of possible relationships between behavioral characteristics and physiological variables which may in turn be traceable to hereditary factors. Research on EEG, autonomic balance, metabolic processes, and biochemical factors illustrate this approach. A lucid demonstration of the process of tracing a psychological condition to genetic factors is provided by the identification and subsequent investigation of phenylpyruvic amentia. In this case, the causal chain from defective gene, through metabolic disorder and consequent cerebral malfunctioning, to feeble-mindedness and other overt symptoms can be described step by step (cf. *32*; *33*, pp. 389–391). Also relevant are the recent researches on neurological and biochemical correlates of schizophrenia (*9*). Owing to inadequate methodological controls, however, most of the findings of the latter studies must be regarded as tentative (*19*).

Prenatal environmental factors provide a third avenue of fruitful investigation. Especially noteworthy is the recent work of Pasamanick and his associates (*27*), which demonstrated a tie-up between socioeconomic level, complications of pregnancy and parturition, and psychological disorders of the offspring. In a series of studies on large samples of whites and Negroes in Baltimore, these investigators showed that various prenatal and paranatal disorders are significantly related to the occurrence of mental defects and psychiatric disorders in the child. An important source of such irregularities in the process of childbearing and birth is to be found in deficiencies of maternal diet and in other conditions associated with low socioeconomic status. An analysis of the data did in fact reveal a much higher frequency of all such medical complications in lower than in higher socioeconomic levels, and a higher frequency among Negroes than among whites.

Direct evidence of the influence of prenatal nutritional factors upon subsequent intellectual development is to be found in a recent, well controlled experiment by Harrell et al. (*16*). The subjects were pregnant women in low-income groups, whose normal diets were generally quite deficient. A dietary supplement was administered to some of these women during pregnancy and lactation, while an equated control group received placebos.

When tested at the ages of three and four years, the offspring of the experimental group obtained a significantly higher mean IQ than did the offspring of the controls.

Mention should also be made of animal experiments on the effects of such factors as prenatal radiation and neonatal asphyxia upon cerebral anomalies as well as upon subsequent behavior development. These experimental studies merge imperceptibly into the fourth approach to be considered, namely, the investigation of the influence of early experience upon the eventual behavioral characteristics of animals. Research in this area has been accumulating at a rapid rate. In 1954, Beach and Jaynes (*8*) surveyed this literature for the *Psychological Bulletin,* listing over 130 references. Several new studies have appeared since that date (e.g., *14, 21, 24, 25, 35*). The variety of factors covered ranges from the type and quantity of available food to the extent of contact with human culture. A large number of experiments have been concerned with various forms of sensory deprivation and with diminished opportunities for motor exercise. Effects have been observed in many kinds of animals and in almost all aspects of behavior, including perceptual responses, motor activity, learning, emotionality, and social reactions.

In their review, Beach and Jaynes pointed out that research in this area has been stimulated by at least four distinct theoretical interests. Some studies were motivated by the traditional concern with the relative contribution of maturation and learning to behavior development. Others were designed in an effort to test certain psychoanalytic theories regarding infantile experiences, as illustrated by studies which limited the feeding responses of young animals. A third relevant influence is to be found in the work of the European biologist Lorenz (*23*) on early social stimulation of birds, and in particular on the special type of learning for which the term "imprinting" has been coined. A relatively large number of recent studies have centered around Hebb's (*17*) theory regarding the importance of early perceptual experiences upon subsequent performance in learning situations. All this research represents a rapidly growing and promising attack on the *modus operandi* of specific environmental factors.

The human counterpart of these animal studies may be found in the comparative investigation of child-rearing practices in different cultures and subcultures. This represents the fifth approach in our list. An outstanding example of such a study is that by Whiting and Child (*38*), published in 1953. Utilizing data on 75 primitive societies from the Cross-Cultural Files of the Yale Institute of Human Relations, these investigators set out to test a number of hypotheses regarding the relationships between child-rearing practices and personality development. This analysis was followed up by field observations in five cultures, the results of which have not yet been reported (cf. *37*).

Within our own culture, similar surveys have been concerned with the diverse psychological environments provided by different social classes (*11*). Of particular interest are the study by Williams and Scott (*39*) on the association between socioeconomic level, permissiveness, and motor development among Negro children, and the exploratory research by Milner (*26*) on the relationship between reading readiness in first-grade children and patterns of parent-child interaction. Milner found that upon school entrance the lower-class child seems to lack chiefly two advantages enjoyed by the middle-class child. The first is described as "a warm positive family atmosphere or adult-relationship pattern which is more and more being recognized as a motivational prerequisite of any kind of adult-controlled learning." The lower-class children in Milner's study perceived adults as predominantly hostile. The second advantage is an extensive opportunity to interact verbally with adults in the family. The latter point is illustrated by parental attitudes toward mealtime conversation, lower-class parents tending to inhibit and discourage such conversation, while middle-class parents encourage it.

Most traditional studies on child-rearing practices have been designed in terms of a psychoanalytic orientation. There is need for more data pertaining to other types of hypotheses. Findings such as those of Milner on opportunities for verbalization and the resulting effects upon reading readiness represent a step in this direction. Another possible source of future data is the application of the intensive observational techniques of psychological ecology developed by Barker and Wright (*7*) to widely diverse socioeconomic groups.

A sixth major approach involves research on the previously cited somatopsychological relationships (*6*). To date, little direct information is available on the precise operation of this class of factors in psychological development. The multiplicity of ways in which physical traits—whether hereditary or environmental in origin—may influence behavior thus offers a relatively unexplored field for future study.

The seventh and final approach to be considered represents an adaptation of traditional twin studies. From the standpoint of the question "How?" there is need for closer coordination between the usual data on twin resemblance and observations of the family interactions of twins. Available data already suggest, for example, that closeness of contact and extent of environmental similarity are greater in the case of monozygotic than in the case of dizygotic twins (cf. *2*). Information on the social reactions of twins toward each other and the specialization of roles is likewise of interest (*2*). Especially useful would be longitudinal studies of twins, beginning in early infancy and following the subjects through school age. The operation of differential environmental pressures, the development of specialized roles, and other environmental influences could thus be more clearly identified and correlated with intellectual and personality changes in the growing twins.

Parenthetically, I should like to add a remark about the traditional applications of the twin method, in which persons in different degrees of hereditary and environmental relationships to each other are simply compared for behavioral similarity. In these studies, attention has been focused principally upon the amount of resemblance of monozygotic as contrasted to dizygotic twins. Yet such a comparison is particularly difficult to interpret because of the many subtle difference in the environmental situations of the two types of twins. A more fruitful comparison would seem to be that between dizygotic twins and siblings, for whom the hereditary similarity is known to be the same. In Kallmann's monumental research on psychiatric disorders among twins (*20*), for example, one of the most convincing bits of evidence for the operation of hereditary factors in schizophrenia is the fact that the degrees of concordance for dizygotic twins and for siblings were practically identical. In contrast, it will be recalled that in intelligence test scores dizygotic twins resemble each other much more closely than do siblings—a finding which reveals the influence of environmental factors in intellectual development.

SUMMARY

The heredity-environment problem is still very much alive. Its viability is assured by the gradual replacement of the questions, "Which one?" and "How much?" by the more basic and appropriate question, "How?" Hereditary influences—as well as environmental factors of an organic nature—vary along a "continuum of indirectness." The more indirect their connection with behavior, the wider will be the range of variation of possible outcomes. One extreme of the continuum of indirectness may be illustrated by brain damage leading to mental deficiency; the other extreme, by physical characteristics associated with social stereotypes. Examples of factors falling at intermediate points include deafness, physical diseases, and motor disorders. Those environmental factors which act directly upon behavior can be ordered along a continuum of breadth or permanence of effect, as exemplified by social class membership, amount of formal schooling, language handicap, and familiarity with specific test items.

Several current lines of research offer promising techniques for exploring the *modus operandi* of hereditary and environmental factors. Outstanding among them are investigations of: (a) hereditary conditions which underlie behavioral differences between selectively bred groups of animals; (b) relations between physiological variables and individual differences in behavior, especially in the case of pathological deviations; (c) role of prenatal physiological factors in behavior development; (d) influence of early experience upon eventual behavioral characteristics; (e) cultural differences in child-rearing practices in relation to intellectual and emotional development; (f) mechanisms of somatopsychological relationships; and (g) psychological development of twins from infancy to maturity, together with observations of their social environment. Such approaches are extremely varied with regard to subjects employed, nature of psychological functions studied, and specific experimental procedures followed. But it is just such heterogeneity of methodology that is demanded by the wide diversity of ways in which hereditary and environmental factors interact in behavior development.

REFERENCES

1. ANASTASI, ANNE. *Psychological testing*. New York: Macmillan, 1954.
2. ANASTASI, ANNE. *Differential psychology*. (3rd ed.) New York: Macmillan, 1958.
3. ANASTASI, ANNE, & CORDOVA, F. A. Some effects of bilingualism upon the intelligence test performance of Puerto Rican children in New York City. *J. educ. Psychol.*, 1953, 44, 1–19.
4. ANASTASI, ANNE, & FOLEY, J. P., JR. A proposed reorientation in the heredity-environment controversy. *Psychol. Rev.*, 1948, 55, 239–249.
5. ARSENIAN, S. Bilingualism in the postwar world. *Psychol. Bull.*, 1945, 42, 65–86.
6. BARKER, R. G., WRIGHT, BEATRICE A., MYERSON, L., & GONICK, MOLLIE R. Adjustment to physical handicap and illness: A survey of the social psychology of physique and disability. *Soc. Sci. Res. Coun. Bull.*, 1953, No. 55 (Rev.).
7. BARKER, R. G., & WRIGHT, H. F. *Midwest and its children: The psychological ecology of an American town*. Evanston, Ill.: Row, Peterson, 1955.
8. BEACH, F. A., & JAYNES, J. Effects of early experience upon the behavior of animals. *Psychol. Bull.*, 1954, 51, 239–263.
9. BRACKBILL, G. A. Studies of brain dysfunction in schizophrenia. *Psychol. Bull.*, 1956, 53, 210–226.
10. DARCY, NATALIE T. A review of the literature on the effects of bilingualism upon the measurement of intelligence. *J. genet. Pychol.*, 1953, 82, 21–57.
11. DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child rearing. *Amer. sociol. Rev.*, 1946, 11, 698–710.
12. DOBZHANSKY, T. The genetic nature of differences among men. In S. Persons (Ed.), *Evolutionary thought in America*. New Haven: Yale Univer. Press, 1950. Pp. 86–155.
13. DOBZHANSKY, T. Heredity, environment, and evolution. *Science*, 1950, 111, 161–166.
14. FORGUS, R. H. The effect of early perceptual learning on the behavioral organization of adult rats. *J. comp. physiol. Psychol.*, 1954, 47, 331–336.
15. HALDANE, J. B. S. *Heredity and politics*. New York: Norton, 1938.

16. HARRELL, RUTH F., WOODYARD, ELLA, & GATES, A. I. *The effect of mothers' diets on the intelligence of the offspring*. New York: Bur. Publ., Teach. Coll., Columbia Univer., 1955.
17. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
18. HIRSCH, J., & TRYON, R. C. Mass screening and reliable individual measurement in the experimental behavior genetics of lower organisms. *Psychol. Bull.*, 1956, 53, 402–410.
19. HORWITT, M. K. Fact and artifact in the biology of schizophrenia. *Science*, 1956, 124, 429–430.
20. KALLMANN, F. J. *Heredity in health and mental disorder; Principles of psychiatric genetics in the light of comparative twin studies*. New York: Norton, 1953.
21. KING, J. A., & GURNEY, NANCY L. Effect of early social experience on adult aggressive behavior in C57BL10 mice. *J. comp. physiol. Psychol.*, 1954, 47, 326–330.
22. LOEVINGER, JANE. On the proportional contributions of differences in nature and in nurture to differences in intelligence. *Psychol. Bull.*, 1943, 40, 725–756.
23. LORENZ, K. Der Kumpan in der Umwelt des Vogels. Der Artgenosse als auslösendes Moment sozialer Verhaltungsweisen. *J. Orn. Lpz.*, 1935, 83, 137–213; 289–413.
24. LUCHINS, A. S., & FORGUS, R. H. The effect of differential postweaning environment on the rigidity of an animal's behavior. *J. genet. Psychol.*, 1955, 86, 51–58.
25. MELZACK, R. The genesis of emotional behavior: An experimental study of the dog. *J. comp. physiol. Psychol.*, 1954, 47, 166–168.
26. MILNER, ESTHER A. A study of the relationships between reading readiness in grade one school children and patterns of parent-child interaction. *Child Develpm.*, 1951, 22, 95–112.
27. PASAMANICK, B., KNOBLOCH, HILDA, & LILIENFELD, A. M. Socioeconomic status and some precursors of neuropsychiatric disorder. *Amer. J. Orthopsychiat.*, 1956, 26, 594–601.
28. SCHWESINGER, GLADYS C. *Heredity and environment*. New York: Macmillan, 1933.
29. SCOTT, J. P., & CHARLES, MARGARET S. Some problems of heredity and social behavior. *J. genet. Psychol.*, 1953, 48, 209–230.
30. SCOTT, J. P., & FULLER, J. L. Research on genetics and social behavior at the Roscoe B. Jackson Memorial Laboratory, 1946–1951—A progress report. *J. Hered.*, 1951, 42, 191–197.
31. SEARLE, L. V. The organization of hereditary maze-brightness and maze-dullness. *Genet. Psychol. Monogr.*, 1949, 39, 279–325.

32. SNYDER, L. H. The genetic approach to human individuality. *Sci. Mon., N. Y.*, 1949, 68, 165–171.
33. SNYDER, L. H., & DAVID, P. R. *The principles of heredity*. (5th ed.) Boston: Heath, 1957.
34. SPOERL, DOROTHY T. Bilinguality and emotional adjustment. *J. abnorm. soc. Psychol.*, 1943, 38, 37–57.
35. THOMPSON, W. R., & MELZACK, R. Early environment. *Sci. Amer.*, 1956, 194 (1), 38–42.
36. TRYON, R. C. Genetic differences in maze-learning ability in rats. *Yearb. nat. Soc. Stud. Educ.*, 1940, 39, Part I, 111–119.
37. WHITING, J. W. M., et al. *Field guide for a study of socialization in five societies*. Cambridge, Mass.: Harvard Univer., 1954 (mimeo.).
38. WHITING, J. W. M., & CHILD, I. L. *Child training and personality: A cross-cultural study*. New Haven: Yale Univer. Press, 1953.
39. WILLIAMS, JUDITH R., & SCOTT, R. B. Growth and development of Negro infants: IV. Motor development and its relationship to child rearing practices in two groups of Negro infants. *Child Develpm.*, 1953, 24, 103–121.
40. WOODWORTH, R. S. Heredity and environment: A critical survey of recently published material on twins and foster children. *Soc. Sci. Res. Coun. Bull.*, 1941, No. 47.

Part VII

INTELLIGENCE TESTING AND ITS CRITICS

16

Lost: Our Intelligence? Why?

QUINN MCNEMAR

One of the effects of the widespread concern with the issues discussed in preceding sections of this book was to leave an impression in the minds of many psychologists, teachers, and educated citizens that science had proved that there was no such thing as general intelligence and that what appeared to be individual differences in intellectual ability were of no importance. In his presidential address to the American Psychological Association, Quinn McNemar launched an attack on some of the sources of this misconception.[1] *It is appropriate that this collection end with a somewhat abridged version of his paper, entitled "Lost: Our Intelligence? Why?",* American Psychologist, *1964,* 19, *871–882. It is reprinted with the permission of the author and the American Psychological Association.*

The Greeks had a word for it, but the Romans had a word with better survival properties. Regardless of the word, what is now called intelligence has been talked about for at least 2,000 years. And as long as 2,000 years before the advent of attempts to measure intelligence, there seems to have been recognition of the fact that individuals differ in intellectual ability.

The earlier attempts at measuring were based on either of two quite distinct conceptions: the Galton-Cattell idea that intellectual ability manifests itself in simple, discrimination functioning, and the Binet notion that cognitive ability reflects itself in more complex functioning. The Binet concept proved to be more fruitful, and by 1925 there was on the market, in addition to various versions of the Binet scale, a flood of group tests of so-called general intelligence.

---

[1] Address of the President to the Seventy-Second Annual Convention of the American Psychological Association, Los Angeles, September 5, 1964.

A few words about definition may be in order. First, it might be claimed that no definition is required because all intelligent people know what intelligence is—it is the thing that the other guy lacks. Second, the fact that tests of general intelligence based on differing definitions tend to intercorrelate about as highly as their respective reliabilities permit indicates that, despite the diversity of definitions, the same function or process is being measured—definitions can be more confusing than enlightening. Third, that confusion might have been anticipated is evident from a recent reexamination of the problem of definition by Miles (1957). This British chappie found himself struggling with the awful fact that the word "definition" itself has 12 definitions. Perhaps the resolution of this problem should be assigned to the newly formed Division of Philosophical Psychology, or maybe the problem should be forgotten since psychologists seem to have lost the concept of general intelligence.

Why has the concept been abandoned? Was it replaced by something else? By something better? Must we admit that the millions who have been tested on general intelligence tests were measured for a nonexistent function? If it is possible that the notion of general intelligence is not lost but merely gone astray, in what corners of what psychological fields should we search for it?

REASONS FOR DISCARDING THE IDEA OF GENERAL INTELLIGENCE

Apparently one reason why concepts are either discarded or modified beyond recognition is that too much is claimed for them. Among the supposed strikes against general intelligence are the following: the earlier false claims about IQ constancy; prediction failures in individual cases; unfounded claims that something innate was being measured by the tests; equally unfounded assertions that nothing but cultural effects were involved; the bugaboo that IQ tests reflect middle-class values; the notion that an IQ standing fosters undesirable expectations regarding school achievement; the idea that IQ differences are incompatible with democracy and lead to educational determinism; and, finally, the great stress on general intelligence caused us to ignore other possible abilities.

This last point leads us right into the problem of factor analysis. Spearman died in battle defending his theory of *g*. Under pressure he reluctantly conceded that factors other than *g* might exist, and he frequently said, in effect, I told you so as long ago as 1906. Actually, Spearman was on the run before the invention of modern factor analysis, but it was not until Thurstone's (1938) first major application of his centroid factor method that Spearman's *g* became, seemingly, nonexistent. Thurstone said, "We have not found the general factor of Spearman" and "We cannot report any general common factor in the battery of fifty-six tests [p. vii]." As anticipated by some, Spearman was not prone to admit defeat. He reworked Thurstone's data and a *g* was found, plus some group factors. He charged that Thurstone's rotational process had submerged the general factor.

American factorists found Thurstone convincing. The description of abilities in terms of seven primaries was an attractive package. The so-called primaries were more amenable to specific definition than the old hodgepodge called general intelligence. Despite the fact that Thurstone was able to replicate his findings on samples from two other populations, thus giving credence to his method and results, there were a couple of events that led to some turbulence in his seven-dimension rarefied atmosphere. The first of these was a minor study, by one of his own students, based on the intercorrelations of 1916 Stanford-Binet items, in which the *g* refused to be rotated out. But rather than admit that this might be some kind of general intelligence, the author renamed it "maturational level." Incidentally, this illustrates the first cardinal Principle of Psychological Progress: *Give new names to old things.*

The second disturber of the neat little set of primaries, sans a *g*, resulted when Thurstone took the next logical step, that of constructing tests to measure the primaries. It was found that the primaries were themselves intercorrelated whereas it had, at the time, been expected and hoped that they would be independent. The Thurstones (1941, p. 26) readily admitted that a general factor was needed to explain the interrelatedness of the primaries. This eventually led to the idea of oblique axes, which axes were regarded as representing the primaries as first-order factors, whereas the general factor pervading the primaries was dubbed a second-order factor. It began to look as though Spearman was being revisited, except for the little matter of labeling: anything called second-order could not possibly be regarded as of much importance. Furthermore, it could always be said that, in the ability domain, it is less difficult to attribute psychological meaningfulness to first-order than to second-order factors, so why pay much attention to the latter? Thus it was easy for most American factorists to drop the concept of general intelligence and to advocate that tests thereof, despite their proven usefulness over the years, should be replaced by tests of the primaries. Hence the emergence of differential aptitude batteries, about which more later.

Meanwhile, our British cousins did not tag along with the factor methods preferred on this side of the Atlantic. After all, it is possible to use factor methods that permit a sizable general factor, if such exists, to emerge as the very first factor. Being first, it is, presto, the most important, as indeed it is as a factor explaining, for the starting battery as a whole, more variance for more tests than attributable to any American-style primary factor. The methods preferred by the British also yield group factors, apt to bear the same name as the primaries, but of attenuated importance. Apparently the British are skeptical of the multitude of ability factors being "discovered" in America. The structure of intellect that requires 120 factors may very well lead the British, and some of the rest of us, to regard our fractionization and fragmentation of ability, into more and more factors of less and less importance, as indicative of scatterbrainedness. This statement presumes that intellectual abilities are brain centered.

In practically all areas of psychological research the demonstration of trivially small minutia is doomed to failure because of random errors. Not so if your technique is factor analysis, despite its being based on the correlation coefficient—that slipperiest of all statistical measures. By some magic, hypotheses are tested without significance tests. This happy situation permits me to announce a Principle of Psychological Regress: *Use statistical techniques that lack inferential power.* This will not inhibit your power of subjective inference and consequently will progress you right back to the good old days when there was no strangling stat or sticky stix to make your insignificant data insignificant.

It may be a long time before we have an ivory tower, strictly scientific resolution of the issue as to whether a scheme involving primary abilities plus a deemphasized *g* is preferable to one involving an emphasized *g* plus group factors. With bigger and better computers we will have bigger, though not necessarily better, factor-analytic studies, but it seems unlikely that such further studies will, in and of themselves, settle the issue under discussion. Until such time as some genius resolves the broader question, so ably discussed by Lee Cronbach in 1957, of the place, if any, of correlational method in a science that aspires to be experimental, we may have to turn to the criterion of social usefulness as a basis for judging whether it is wise to discard general intelligence. Like it or not, much of our heritage in this area is that earlier workers, from Binet on, had as their motivation the solution of social problems, and currently many in the area have a similar motivation.

THE BEARING OF SOCIAL USEFULNESS

In practice, if you believe that the concept of general intelligence has outlived its usefulness, you may choose from among several differential, or multiple, aptitude batteries, which will provide measures of some of the so-called primary mental abilities. If you happen to believe that there is something to general ability, you can find tests to use. The novice looking for these latter tests may have to alert himself to the first Principle of Psychological Progress—the test labels may have changed from "general intelligence" to "general classification" or "scholastic aptitude." If you enjoy riding the fence, you might become a devotee of the practice of the College Board, and others, and measure just two abilities: Verbal and Quantitative.

This is certainly not the place to review the voluminous literature that amply demonstrates the practical utility of tests of general intelligence. Nor is it the place to catalog the misuses of the Stanford-Binet for purposes which Terman never claimed for it, or the misuses of the Wechsler scales for purposes which Wechsler *has* claimed for his scales. Neither the Binet nor the Wechsler provides a factorially pure, unidimensional measure of a *g*. The current Stanford-Binet was in reality constructed too early to benefit from the implication of factor analysis for test purity, whereas the Wechsler scales were based on the impossible premise that 10 or 11 subtests can simultaneously provide diagnostic subscores and a meaningful total score. Of the many group tests that appeared between 1920 and 1945 it can be said that few, if any, provide unidimensional measures of general intelligence. The chief difficulty is that most of them lead to a total score based on a mixture of verbal and mathematical material. Thus, with two main sources of variance, marked qualitative differences can exist for quantitatively similar total scores. The College Board-Educational Testing Service people have justifiably refrained from giving a total score involving verbal plus math, but there are those who question the usefulness of the Board's math score and there are those who criticize the Educational Testing Service for failing to change over to a differential aptitude battery. . . .

Just how successful have the multitest batteries been? Since by far the most extensive social use of tests has been, and continues to be, in the schools, let us look at the evidence of validity studies therein. As indicated previously, little is known about the predictive usefulness of some of the seven batteries discussed above. The DAT of the Psychological Corporation is the only battery for which adequate predictive (and concurrent) validity data, derived from school sources, are available. It is also the battery that has fared best in the hands of the test reviewers; therefore if we allow the case for differential batteries to rest thereon, we will be looking at the best. So, what is the story?

Recall that the hoped-for advantage of a multitest battery over the old-fashioned general intelligence test was that it would have greater predictive power, a power which could manifest itself in higher validity coefficients for specific subject matter and, perhaps, for overall achievement. It was hoped that such a battery would be truly differential in that particular factors (or subtests) would correlate higher with achievement in some areas than in other areas. Presumably each factor (or subtest) should have unique usefulness. If a battery were truly differential, it would be a boon to school guidance personnel.

Now the manual of the DAT of the Psychological Corporation contains a staggering total of 4,096, yes, I counted 'em, validity coefficients. With such a large pool to draw from, one could by gracious selection "show" that the DAT is the answer to the prayer of every counselor, male or female, or by malicious selection one could "prove" that the DAT is far worse than any test ever published. The validity coefficients range all the way down to −.37, which is presumably a chance deviation downward from 0, and all the way up to .90, which is likely not a chance deviation downward from unity. But ranges tell us nothing. After a careful perusal of the 4,096 correlations, it seems safe to summarize DAT validities as follows:

1. Verbal Reasoning (analogies to most of you) is the best single predictor; Language Usage, as represented by a sentence test dealing with grammar and word usage, and admittedly more achievement than aptitude, is a close second.

2. Numerical Ability, as measured by a test of simple arithmetic operations, designed to tap arithmetic reasoning without the usual verbal component, is the best predictor of achievement in school mathematics. It does not, however, correlate as well with grades in science as does Verbal Reasoning.

3. Aside from the Numerical Ability test, the only other test that shows differential power as a predictor is the Spelling test —if you cannot spell you may have trouble learning shorthand.

4. The remaining five tests in the battery simply fail to show compelling evidence that they are good in the differential predictive sense. For the Mechanical Reasoning and the Clerical Speed and Accuracy test this may be understandable in that little of the school curricula for Grades 8 through 12 requires such abilities, but one would expect that Abstract Reasoning and Space Relations would fare better than they seem to.

Such data as we have been able to locate for the other six multitest batteries tend to support these findings on the DAT. Aside from tests of numerical ability having differential value for predicting school grades in math, it seems safe to conclude that the worth of the multitest batteries as differential predictors of achievement in school has not been demonstrated. Incidentally, the fact that the Verbal and Numerical tests stand out as the only two useful predictors tends to provide some support for the Educational Testing Service-College Board practice of providing scores for just these two abilities.

And now we come to a very disturbing aspect of the situation. Those who have constructed and marketed multiple aptitude batteries, and advocated that they be used instead of tests of general intelligence, seem never to have bothered to demonstrate whether or not multitest batteries provide better predictions than the old-fashioned scale of general intelligence. Be it noted that we are not discussing experimental editions of tests. Some may say that insofar as a test publisher provides validity data for a new battery it is not necessary to show that the validities are, for the given school condition, better than those of other tests. With this one can agree, but only in case no claims are made, explicitly or implicitly, regarding superior merits for the new battery.

It is far from clear that tests of general intelligence have been outmoded by the multitest batteries as the more useful predictors of school achievement. Indeed, one can use the vast accumulation of data on the validity of the Psychological Corporation's DAT to show that better predictions are possible via old-fashioned general intelligence tests. Consider the fact that a combination of the tests Verbal Reasoning (analogies) and Numerical Ability would be, in terms of content, very similar to many group tests of general intelligence. Consider also that an equally weighted combination of these two tests correlates in the mid-.80s with the Otis S-A, Higher Form. Then, when you turn to a careful study of the empirical validities, as reported in the DAT manual, you will not be surprised at the outcome of the application of a little arithmetic, which leads to the definite conclusion that a simple unweighted combination of the Verbal Reasoning and Numerical Ability tests predicts as well as or, in most instances, better than any subtest taken singly, or in the differential sense.

The manual for the DAT contains the following statement (Bennett, Seashore, & Wesman, 1952):

> Apparently the *Verbal Reasoning* and *Numerical Ability* tests can serve most purposes for which a general mental ability test is usually given in addition to providing differential clues useful to the counselor. Hence, the use of the so-called intelligence test is apparently unnecessary where the *Differential Aptitude Tests* are being used [p. 71].

Anyone who disagrees with this quotation could, with better justification, say that an intelligence test can serve nearly all, if not all, the purposes for which a multiple aptitude battery is given in the schools because the former, in general, is a better predictor and because, as we saw earlier, the differential clues are too fragmentary to be of use to the counselor. And there is a bonus: one classroom period of testing, compared to six periods. A second bonus: much less costly. A third bonus: fewer scores to confuse the already confused minds of most school counselors.

Thus, we come to the conclusion that general intelligence has not been lost in the trend to test more and more abilities; it was merely misplaced by a misplaced emphasis on a hope that a lot of us, including the speaker, once entertained, a hope that in turn was based on a misplaced faith in factor analysis: *the* hope that factors, when and if measured, would find great usefulness in the affairs of society. By the criterion of social usefulness, the multiple aptitude batteries have been found wanting. Now, I have no desire to furnish ammunition for those test critics who would have us stop all testing merely because they find a trivially faulty item in a standardized test. At a time when there is shouting about the tyranny of the testers and the brass of the brain watchers, at a time when school people are showing resentment at the disruption caused by too many national testing programs, at a time when federal and state legislators are all too willing to write legislation that places restrictions on the use of tests, and at a time when both majorities and minorities are being denied the benefits of test-based guidance because certain wellintentioned persons fail to realize that scores for the underprivileged minorities are useful indices of *immediate,* or present, functioning—at a time when all these and other forces are operating to throw out the tests, it is high time for the profession to establish a bureau of standards to test the tests instead of coasting down a road that is tinged with some of the trappings of Madison Avenue. Better to have informed internal control than ignorant, hostile, external control.

INTELLIGENCE ELSEWHERE?

Aside from the near loss of the idea that progress in school may depend on general intelligence, one wonders whether intelligence has come to be regarded as unimportant in other areas.

Any of you who have money invested in stocks and wish some reassurance regarding the intelligence level of business and industry managers should read Edwin Ghiselli's (1963) Bingham Lecture. His summary of his own work indicates that the average intelligence of those in the upper and middle management levels falls at the ninety-sixth percentile of the population. Thomas Harrell (1961) came to a similar conclusion. Furthermore, management level is correlated with intelligence—you can be too dumb to succeed as a manager. Also you can be too bright to be a managerial success! Now it must be admitted that little, if anything, is known about whether management success might be better predicted by measures of factor-analytic defined abilities. On this you are free to guess—most of you will have already guessed my guess.

A one-by-one cataloguing of what we know or do not know about what abilities contribute to success within various occupational and professional groups would merely add to the dullness of this presentation, so let us turn to some of the more esoteric fields of psychology to see whether the concept of general intelligence has or has had any relevance. One such field, and a very broad one, is creativity. Anyone who peeks over the fence into this field is apt to be astonished at the visible chaos. The definition of creativity is confounded by the diversity of subareas within the field, the criterion problems are far from licked, and so little is known about the creative process that measuring instruments are, seemingly, chosen on a trial-and-error basis. . . .

[*McNemar summarizes here a number of studies purporting to show that intelligence plays only a minor role in creative achievement and that the correlation between tested intelligence and creativity is very low. He concludes that such studies do not prove their points because of defects in their design.*]

Next we turn to a few studies of creativity which cannot be criticized because of restriction of range on the criteria—these studies simply avoid this problem by never having actual criterion information. The approach is to claim that certain tests, which typically are scored for novel responses or novel solutions to problems, *are* measures of creativity, with no evidence whatsoever that the tests have predictive validity for nontest, real-life creative performance. This bit of ignorance does not prove to be a handicap to those who think that creativity can be studied without the nuisance of obtaining criterion measures. We reluctantly accept the test-based criteria solely for the sake of seeing what happens to general intelligence as a part of the picture. Time permits only three examples.

We first note that general intelligence has not manifested itself as a correlate of so-called creativity tests in the factor-analytic studies of creativity. The explanation for this is easily found—no measures of general intelligence are used in these studies. When discussing his plans for studying creativity, a certain author (Guilford, 1950) said that "we must look well beyond the boundaries of the IQ if we are to fathom the domain of creativity [p. 448]." He went on to say, the conception "that creative talent is to be accounted for in terms of high intelligence or IQ . . . is not only inadequate but has been largely responsible for lack of progress in the understanding of creative people [p. 454]." With a part of this one can agree, but does it follow that one should prejudge the role of general intelligence as a source of variance in creativity tests or factors derived therefrom? Does the failure to include an IQ test help one learn the extent to which one must go beyond the boundaries of the IQ to fathom creativity? Apparently the author, although willing to predict that the correlations between IQ and the many types of creativity tests "are only moderate or low," was unwilling to include an IQ test for the sake of finding out. However, negation by omission is not very convincing.

That at least one test bearing the label "creativity" is correlated more than moderately with IQ is evidenced by the value of .67 (average for boys and girls) for the carefully chosen sample of 15-year-olds in Project Talent (Shaycoft, Dailey, Orr, Neyman, & Sherman, 1963). This sample-stable *r* (based on a total *N* of 7,648) becomes .80 when corrected for attenuation.

In a recent extensive study (Getzels & Jackson, 1962), already extensively criticized, creativity is defined as the sum of scores on five tests (median intercorrelation of only .28). Although the investigators use the sum score for most of their analyses, they do not bother to report the correlation of creativity, so defined, with IQ. From the published report I have ascertained (via the correlation-of-sums formula) that creativity and IQ correlate to the extent of .40 for the total of 533 cases. Now this *r* of .40 has been greatly attenuated because of three things: first, the usual measurement errors; second, the cases were highly selected on IQ (mean of 132); third, the IQs are a mixture from the Stanford-Binet, Henmon-Nelson, and Wechsler Intelligence Scale for Children (the use of regression-estimated Binet IQs from the other two scales aggravates rather than improves the mixture). We deduce that intelligence and creativity tests used here have far more common variance than the authors believe.

Much is made of the finding that the creativity tests tended to correlate higher than did IQ with verbal-content school achievements. Again the IQ comes in for an unfair drubbing because of the same mixture of IQ scores and, what is more pertinent, because of explicit selective curtailment on the IQ variable and only incidental selection on the creativity variable.

Of more importance to the present paper is the analysis, by these same authors, based on a high IQ group and a high creative group, these groups being selected as the top 20% for each variable but excluding those who were in the top 20% on both variables. These two selected groups were then contrasted on total school achievement (and a host of other variables that are of no interest here). The mean IQ for the high IQ group was

150 whereas the mean IQ for the high creative group was 127, yet the achievement means of the two groups were "unexpectedly" equally superior to the school population mean despite the 23-point difference in mean IQ. The authors say that it "*is* quite surprising" that the high creativity group achieved so well. From this it is concluded that the "creative instruments account for a significant portion of the variance in school achievement [p. 24]," and the subsequent argument implies that creativity is more important for ordinary school achievement than is the IQ. Now anyone who is at all familiar with a three-variate problem will not be "unexpectedly" surprised at the foregoing results—indeed, if the authors had bothered to give the three basic correlations among the three variables (IQ, creativity, and total school achievement) for the entire group, any person versed in simple multivariate analysis could deduce the results. Furthermore, he could deduce a further result (and this one has been overlooked by the critics) which might be unpleasantly surprising to the thesis of these authors: namely, the high IQ and the high creative groups did equally well in school achievement despite an unreported difference in mean creativity that is of the same order as the much stressed difference in IQ.[2] Utilizing the half-blind logic of the authors, one can say that creative ability is not as important as IQ for school achievement—just the opposite of their position.

Now the fact that seven of nine replications of this study confirm the original findings merely indicates that repetition of the same faulty design and false logic will lead to the same false conclusions. The design being used is such that, if two variables are equally correlated with a third, the conclusion will be reached that the two are actually unequally correlated with the third. This is the neatest trick of the decade for supplying educationists with an antidote for the IQ virus. I cannot refrain from saying at this point that, although discouraged, I am still hopeful that people who do statistical studies will first learn a modicum of elementary statistics!

Time does not permit a discussion of other studies in which creativity is defined in terms of test performance instead of being based on actual creativity of the sort prized by society. In summary of this brief on creativity studies, I would like to offer a few dogmatic-sounding observations. First, one need not be surprised at the fact that so-called creativity tests do not yield high correlations with IQ tests—but the correlations are generally far higher than those found in typical studies with range restrictions. I would anticipate that for normalized scores, the uncurtailed scatters for IQ versus creativity tests will be bivariate normal. Second, if we have honest to goodness criterion measures of literary or architectural or scientific creativity, the scatter diagram between IQ and such creativity (not normalized, since it makes sense to expect a skewed distribution for actual creativity) will be triangular in shape for unselected cases. That is, at the high IQ levels there will be a very wide range of creativity, whereas as we go down to average IQ, and on down to lower levels, the scatter for creativity will be less and less. Having a high IQ is not a guarantee of being creative; having a low IQ means creativity is impossible. Third, it remains to be seen whether or not the so-called creativity tests and/or factors derived therefrom have appreciable value as predictors of actual creative performance. Such tests may or may not yield better predictions than a test of general intelligence. Fourth, as far as I am concerned, to claim factorial validity for creativity tests, along with definitions of creativity in terms of tests, is an unwarranted avoidance of the fundamental problem of validity.

---

[2] Since this was written, the replication study of Yamamoto (1964) gives data that corroborate this deduction.

The recently renewed interest in "gifted" children, along with the flurry of creativity studies, has led to a reexamination of methods for identifying the gifted. It has long been recognized that identification in terms of high IQ is too narrow—those gifted in such areas as art and music would be overlooked. The argument against the IQ is now (Torrance, 1962) being reinforced by the claim that the selection of the top 20% on IQ would mean the exclusion of 70% of the top 20% on tested creativity. This startling statistic, which implies a correlation of only .24 between IQ and creativity, is being used to advocate the use of creativity tests for identifying the gifted. Be it noted that these creativity tests will also miss those gifted in art and music.

We are being told that it is important "to identify creative talent early in life," hence you need not be surprised that the search goes down to the kindergarten level, with claims of successful identification. The creativity tests are presumed to be better for this purpose than the IQ tests because of the failure of the IQ to be constant, an argument that completely overlooks the fact that the IQ does have some constancy whereas absolutely nothing is known about the stability of standings on creativity tests. The IQ tests, known to be imperfectly valid as predictors of outstanding achievement in life, are to be replaced by the creativity tests, known to be of unknown validity as predictors. Anyway, progress, defined as change, is in the offing.

The IQ is being linked with *learning* as an outmoded educational objective; the new objective involves an emphasis on *thinking*. Somehow or other creativity, not general intelligence, is being associated with thinking. The horrible idea of underachievers and overachievers, in terms of expectancies based on the IQ, will be abolished. But no thought is given to the fact that the use of creativity tests will simply define a new crop of under- and overachievers.

In an apparent zeal to rid us of general intelligence, it is argued that measured creativity is significantly related to ordinary school achievement. Maybe so, but never, never does one find complete data reported as to the relative sizes of validity coefficients. And, as we have seen, the technique being used will show that equal coefficients are unequal. Why not the full facts, free of fantasy?

An additional difficulty is not being faced by those who would replace IQ tests by creativity tests, or creative-thinking tests. The factor-analytic studies indicate either no, or a trivially small, general creativity factor in these tests, yet these self-characterized "bold, adventurous" reformers (see Torrance, 1963) do not hesitate to advocate a total score which is nearly devoid of meaning. Changing the curriculum to the teaching of creativity and creative thinking will not overcome this measurement difficulty. Again, I express the hope that the IQ is replaced by something better rather than by something worse.

There are other areas, such as reasoning, problem solving, and concept formation, in which one might expect to find some consideration of intelligence as an aspect. One might also expect that investigators of thinking would have something to say about individual differences in thinking being dependent upon intelligence, but for some unintelligent reason these people seem never to mention intelligence. Surely, it cannot be inferred that thinking about thinking does not involve intelligence!

IN CONCLUSION

It has been the thesis of this paper that the concept of general intelligence, despite being maligned by a few, regarded as a second-order function by some, and discarded or ignored by others, still has a rightful place in the science of psychology and in the practical affairs of man. It has not been argued that the nature of general intelligence is well understood. Much, however, has been written about its nature. Over 40 years ago (Intelligence, 1921a, 1921b), an editor secured and published the reasoned views of 13 well-known test psychologists. Later, Spearman set forth his speculations about the nature of *g*. Prior to these, Binet had, of course, given much thought to the problem.

More recent discussions exist. Hebb (1949) has considered the problem from the viewpoint of neurology and brain functioning. Cyril Burt (1955), always a vociferous defender of the concept of general intelligence, has reviewed the evidence for a *g* and restated the idea, dreadful to some, that intelligence is innate. Perhaps it was inevitable that Raymond Cattell (1963), who has camped with the general intelligence contingent, should gaze into his crystal *n*-dimensional factor ball and find evidence for crystallized as opposed to fluid general intelligence. Joseph McVicker Hunt's (1961) book on *Intelligence and Experience* is in large part devoted to questions pertaining to the nature of intelligence.

By far the most provocative recent discussion that I have encountered is the closely reasoned 44-page paper by Keith Hayes (1962). He puts forth a motivational-experiential theory of intelligence. In essence, he presumes that there are hereditary differences in motivation. "Experience-producing drives" and environmental differences produce differences in experience, which in turn, by way of learning, lead to differences in ability. Therefore, differences called intellectual are nothing more than acquired abilities. I think that Hayes has ignored the possibility of individual differences in learning ability, but if such a formulation leads to experimental manipulation of variables, we may eventually make progress in an area that has too long been dominated by ever increasing fractionization by factor analysis, with little thought as to how the fractured parts get put together into a functioning whole.

Abilities, or capacities, or aptitudes, or intellectual skills, or whatever you choose to call them, are measured in terms of response products to standardized stimulus situations. The *stimulus* is presented to an *organism* which by some *process* comes up with a *response;* thus any attempt to theorize and/or study intellect in terms of a simple stimulus-response (S-R) paradigm seems doomed to failure unless drastically modified and complicated by the insertion of O for organism and P for process.

There have been thousands of researches on the multitudinous variations from organism to organism, and the results fill books on individual differences. These studies can be roughly classified into two types. First, those that ascertain the intercorrelations among scaled response products to various stimulus situations, known as tests, have to do with the structure of intellect; and whether the resulting factors are anything more than dimensions for describing individual differences need not concern us here. The second type of study seeks the nontest correlates of test performance, and whether or not any of the found correlates can be regarded as explaining individual differences is not of interest here. Both types of studies certainly force one to stress the overwhelming diversity exhibited among the organisms.

But these studies of individual differences never come to grips with the *process,* or operation, by which a given organism achieves an intellectual response. Indeed, it is difficult to see how the available individual difference data can be used even as a starting point for generating a theory as to the process nature of general intelligence or of any other specified ability.

As a basis for a little speculation, let us conceive of a highly hypothetical situation in which the two members of a pair of identical twins, with identical experiences, find themselves cast up on an uninhabited tropical island. Let us assume that they are at the supergenius level, far beyond that of your favorite man of genius. Let us also assume that, though highly educated in the sciences, they have been fortunate enough to have had zero exposure to psychology. In addition, we presume that, being highly involved and abstracted in the pursuit of science, they have never noticed what we call individual differences in abilities.

A quick exploration of the island assures them that food is plentiful, that shelter is available, and that clothing is not a necessity. To allay the boredom that they foresee as an eternity in this laborless heaven, they decide to spend their time in the further pursuit of science, but the lack of the wherewithal for constructing gadgets rules out any research in the physical sciences. Having had a college course in Bugs and Bites they proceed to study the life of the island's insects, then the habits of the birds, and the antics of a couple of monkeys. The manner in which the monkeys adjust to the environment leads them to set up some trial situations for more systematic observation. Needless to say, the monkeys show evidence of what we call learning and what we call problem solving.

Eventually they decide that attempting to outwit each other might be more fun than being outwitted by the monkeys, so they begin to cook up and use games and problems for this purpose. This activity leads each to speculate and introspect about how problems are invented and how solved. Then by cleverly designed experiments, preceded of course by theory, they set forth highly developed laws and principles about what we call reasoning and problem solving. Incidentally, they switch back and forth between the roles of experimenter and subject, there being no college sophomores available. They continue for years the study of their own mental operations, constantly on the alert for new phenomena to investigate.

And now with apologies to the ancient Greeks, who did have some ideas along these lines, we leave with you the 64-million drachma question: Will our two identical supergeniuses, being totally unaware of individual differences, ever hit upon and develop a concept of intelligence?

S.C.E.R.T., West Bengal Library

REFERENCES

BENNETT, G. K., SEASHORE, H. G., & WESMAN, A. G. *Differential Aptitude Tests, manual.* (2nd ed.) New York: Psychological Corporation, 1952.

BURT, C. L. The evidence for the concept of intelligence. *Brit. J. educ. Psychol.*, 1955, 25, 158–177.

CATTELL, R. B. Theory of fluid and crystallized intelligence: A critical experiment. *J. educ. Psychol.*, 1963, 54, 1–22.

CRONBACH, L. J. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, 12, 671–684.

GETZELS, J. W., & JACKSON, P. W. *Creativity and intelligence.* New York: Wiley, 1962.

GHISELLI, E. E. Managerial talent. *Amer. Psychologist*, 1963, 18, 631–642.

GUILFORD, J. P. Creativity. *Amer. Psychologist*, 1950, 5, 444–454.

HARRELL, T. H. *Manager's performance and personality.* Cincinnati, O.: South-Western, 1961.

HAYES, K. J. Genes, drives, and intellect. *Psychol. Rep.*, 1962, 10, 299–342.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HUNT, J. McV. *Intelligence and experience.* New York: Ronald Press, 1961.

Intelligence and its measurement: A symposium. *J. educ. Psychol.*, 1921, 12, 123–147. (a)

Intelligence and its measurement: A symposium. *J. educ. Psychol.*, 1921, 12, 195–216. (b)

MILES, T. R. On defining intelligence. *Brit. J. educ. Psychol.*, 1957, 27, 153–165.

SEGEL, D. The multiple aptitude tests. *Personnel guid. J.*, 1957, 35, 424–432.

SHAYCOFT, M. F., DAILEY, J. T., ORR, D. B., NEYMAN, C. A., JR., & SHERMAN, S. E. Project Talent: Studies of a complete age group—age 15. Pittsburgh: University of Pittsburgh, 1963. (Mimeo.)

THURSTONE, L. L. *Primary mental abilities.* Chicago: Univer. Chicago Press, 1938.

THURSTONE, L. L., & THURSTONE, T. G. *Factorial studies of intelligence.* Chicago: Univer. Chicago Press, 1941.

TORRANCE, E. P. *Guiding creative talent.* Englewood Cliffs, N.J.: Prentice-Hall, 1962.

TORRANCE, E. P. *Education and the creative potential.* Minneapolis: Univer. Minnesota Press, 1963.

YAMAMOTO, K. Role of creative thinking and intelligence in high school achievement. *Psychol. Rec.*, 1964, 14, 783–789.

Form No. 3.

PSY, RES.L-1

Bureau of Educational & Psychological Research Library.

The book is to be returned within the date stamped last.

....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....

WBGP-59/60-5119C-5M

E.R.T., West
Library

VAN NOSTRAND INSIGHT BOOKS

1 THE CRISIS IN PSYCHIATRY AND RELIGION–Mowrer
* 2 INSTINCT–Birney and Teevan
* 3 REINFORCEMENT–Birney and Teevan
* 4 COLOR VISION–Teevan and Birney
5 TOWARD A PSYCHOLOGY OF BEING, *2nd ed.*–Maslow
* 6 MEASURING HUMAN MOTIVATION–Birney and Teevan
* 7 EMOTION–Candland
* 8 PROGRAMMED LEARNING–Smith and Moore
* 9 TRANSFER OF LEARNING–Grose and Birney
10 CONT ARY RESEARCH IN LEARNING–Braun
11 RESE ATHOLOGY–Quay
12 PSYC
13 RACE
*14 PRIM
15 THE
16 THE
17 THE
18 THE
*19 THE
*20 THE
PSY
21 TH
22 AN
23 CO
24 PE
*25 SO
*26 PS
*27 F
28 T
29 U
*30 C
C
*31
32
33
*34
35
36 ff
*37
38

Form No. 4

BOOK CARD

Coll. No.............. Accn. No......

Author..............................

Title...............................

....

Date.	Issued to	Returned on
....	....	....
....		

PSYCHOLOGY

151
TYL

About this book:

Teachers, college admissions officers, school counse [illegible] personnel managers – these are only a segment of the [illegible] frequently called upon to make decisions based (at leas [illegible] on the results and interpretation of intelligence testing. W [illegible] an awareness of what is (and what is not) being measured, they are likely to place either too much or too little reliance on the information such tests give them. Unfortunately, many of the most illuminating discussions and significant research reports about intelligence and the problem that one faces in attempting to measure it are in journals and books not easily accessible to the majority of people. This collection of articles – each an important contribution to the understanding of intelligence measurement – forms a unique reader, at an unsophisticated technical level, that will appeal to all interested in the subject.

VAN NOSTRAND INSIGHT BOOKS

Insight Books are devoted to filling a vital need in the efficient study of human behavior. By making available, in paperback, essential materials that are frequently inaccessible or often neglected in a more expensive form, the Insight Series renders a valuable service to students, teachers, and interested laymen. Drawn from psychology and its allied fields, the bulk of the books published will deal with the following categories: 1) studies of enduring problems in psychology through original sources; 2) selections from their own works by contemporary behavioral scientists; and 3) authoritative discussions of timely, even controversial, issues.

DAVID C. McCLELLAND is general editor of Insight Books. Currently Professor of Psychology in the Department of Social Relations at Harvard University, he is also an established author in experimental and theoretical psychology. Among his recent books is *The Achieving Society*, published by Van Nostrand.

Along with Dr. McClelland, Dr. Robert C. Birney of Hampshire College and Dr. Richard C. Teevan of Bucknell University are jointly responsible for editing and developing those books in this series devoted to Enduring Problems in Psychology.

VAN NOSTRAND REINHOLD CO.